8.306
3.9.70

# Journal of Experimental Education

……

A QUARTERLY
Edited by
A. S. BARR

With the Cooperation of
JACOB O. BACH
ARTHUR T. JERSILD
H. H. REMMERS
J. WAYNE WRIGHTSTONE

BUREAU OF EDNL. & PSYL. RESEARCH : : LIBRARY.

……

VOLUME XXVIII
September 1959 - June 1960

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

3.9.70
J - 306

# TABLE OF CONTENTS

## VOLUME XXVIII

SEPTEMBER 1959

The Development of Scales of Attitudinal Dimensions: T. Bentley Edwards and Alan B. Wilson . . . . . . . . . . 3

Studies in Social Perception: Word Productivity: Frank J. Estvan . . . . . . . . . . 37

The Relationship Between Handwriting Pressure and Legibility of Handwriting in Children and Adolescents: Theodore L. Harris and G. Lawrence Rarick . . . . . . . . . . 65

The Russell Sage Social Relations Test—A Technique for Measuring Group Problem Solving Skills in Elementary School Children: Dora E. Damrin . . . . . . . . . . 85

DECEMBER 1959

Development of an Instrument for Studying Verbal Behaviors in a Secondary School Mathematics Classroom: E. Muriel J. Wright . . . . . . . . . . 103

The Semantic Method for Testing Vocabulary: Mildred M. Berwick . . . . . . . . . . 123

The Development of the Early School Personality Questionnaire: Richard W. Coan and Raymond B. Cattell . . . . . . . . . . 143

The Relative Validity of Verbal vs. Illustration Items in Achievement Type Tests for Low Verbal Ability Personnel: Peter F. Merenda and Charles J. Macaluso . . . . . . . . . . 153

Aural Perceptions of Authoritarians in Different Cultural Situations: Jack R. Frymier . . . . . . . . . . 163

Group Structure, Anxiety, and Problem-Solving Efficiency: Frank W. Banghart . . . . . . . . . . 171

MARCH 1960

Multiple factor Analysis of Academic Achievement—A Comparitive Study of Full-Blooded Indian and White Children: Omer John Rupiper . . . . . . . . . . 177

Attitude of Kindergarten Children Toward Specially Prepared Story Materials: Bill Snider . . . . . . . . . . 207

Analysis of MMPI Data for Classification Purposes by Multivariate Statistical Techniques: Peter P. Rempel . . . . . . . . . . 219

On Programming the Integral Rth Root of a Number on Friden Standard and Square Root Desk Calculators: William J. Moonan............. 229

Must All Tests be Multi-Factor Batteries? Edwin F. Rosinski............... 235

An Evaluation of the Laboratory in a College General Biology Course: Douglas M. Dearden.......................................... 241

Application of Matrices in the Analysis of Sociometric Data: John X. Jamrich................................................ 249

A Table to Facilitate the Calculation of Normalized T-Scores for Small Samples ($N > 5 < 51$): Peter F. Merenda ......................... 253

Estimation of Test Reliability by Analysis of Variance Technique: Cyril J. Hoyt and P R. Krishnaiah.................................. 257

On Teaching F Instead of t: Norman H. Anderson.......................... 261

Computation of Variance Accounted for in Multiple Correlation: Clinton I. Chase.............................................. 265

On a Test of Whether Two Sets of Observations Belong to a Certain Class of Distributions When the Random Variables are not Independent: H. P. Kuang ........................................ 267

JUNE 1960

The Perception of Music Symbols in Music Reading by Normal Children and by Children Gifted Musically: Robert G. Petzold............. 271

A Study of Relationships Between Importance of a School Superintendent's Functions and His Practices, Based on Authoritative Opinion: Raymond J. Young and Beryl Dillman..................... 321

A Comparative Study of Student-Teaching Practices in Thirty-Eight Midwest Institutions: Gail M. Inlow............................... 337

# The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1959 by
DEMBAR PUBLICATIONS, INC.
Madison, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 1, September 1959)

# THE DEVELOPMENT OF SCALES OF ATTITUDINAL DIMENSIONS*

T. BENTLEY EDWARDS and ALAN B. WILSON
University of California
Berkeley, California

IT HAS BEEN suggested[1]** that the attitudes of students toward school subjects might be illuminated by analyzing the habitual orientation of students toward decision or choice situations according to two dimensions. One dimension is provided by considering the degree of deliberation involved in making a choice: the preference for ends suggested by abstract, relational consideration, as contrasted with the appreciation of immediate, proximate ends; the other dimension is provided by the relative preference for social as opposed to non-social objects.

The intersection of these two dimensions yields four vectors: 1) "theoretic," the deliberative, analytic orientation toward the non-social environment, characterized by interest in the natural sciences and in mathematics; 2) "prudent," the deliberative orientation toward the social environment, characterized by interest in the behavioral sciences; 3) "aesthetic," the appreciation of the quality of the directly given non-social environment; and 4) "immediate," the preference for proximate social ends.[2]

The ranges between the poles of these four vectors provide six continua for the assessment of the structure of attitudes: 1) prudent vs. theoretic, 2) prudent vs. immediate, 3) prudent vs. aesthetic, 4) theoretic vs. immediate, 5) theoretic vs. aesthetic, and 6) aesthetic vs. immediate. These abstractions are summarized by the design in Figure 1.

If an individual who selects relatively proximate ends in a variety of situations may be reliably expected to select proximate ends in a future choice situation, characterizing him as "immediate" (or "aesthetic," in a non-social context) would be of value for the prediction of behavior. More important for educators, however, is the possibility that analysis of decision behavior with the aid of these dimensions may provide a useful theoretical perspective for the evaluation of curricular objectives. An instrument which "measures" an individual's position on the continua could be used to evaluate the effects of curricular modification.

In this paper we will describe efforts to construct six scales in such a fashion that scale scores obtained by individuals will be valid indexes of their relative positions on the six continua.

## Construction of the Scales

A large number of short propositions expressing preference for one of two alternatives were written. The content of each alternative was designed to represent one pole of one continuum, and was paired, in the proposition, with an alternative relatively approximating the opposite pole, e.g., prudent vs. theoretic. From this fund of propositions, thirty were selected for each of the six scales, making 180 items in all for the six scales. The following criteria governed the editing and selection of items:

1. Theoretical relevance of content
2. Clarity of meaning
3. Appropriateness of vocabulary and content for high school students
4. Balance of proportion of "positive" vs. "negative" items (one pole of each continuum was arbitrarily designated positive or favorable; hence disagreeing with a negative item would be a positive response)
5. Avoidance of dyslogistic phraseology or alternatives counter to cultural universals.
6. Avoidance of factual items, responses to which might not reflect attitudes.

These criteria are comparable to those suggested in current literature[3] on attitude scales except that each item contains two ideas, usually explicitly stated, though sometimes implicit, to be compared. It would seem, however, that this greater grammatical complexity is offset by the simplicity of decision in such comparison.

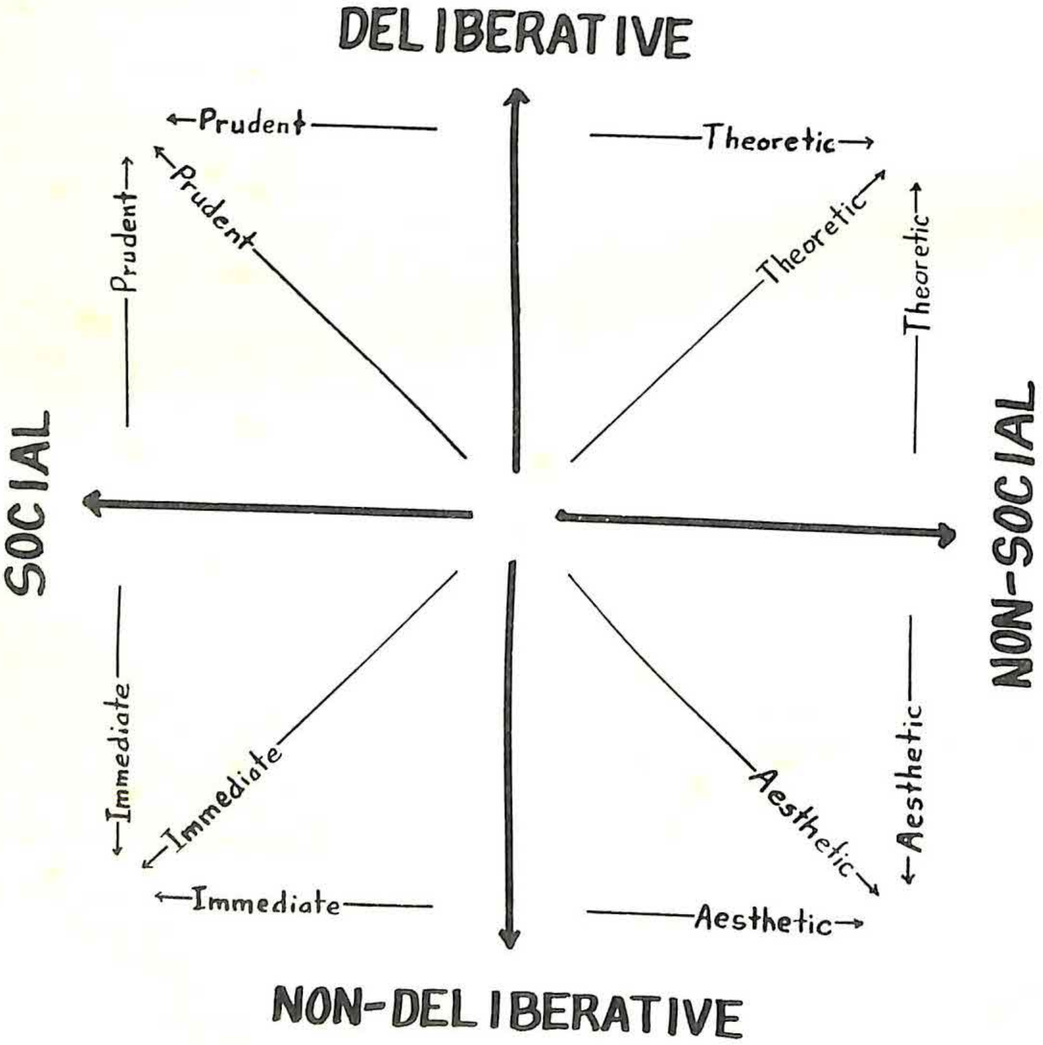
---

* The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare.

**All other footnotes will be found at end of article.

FIGURE 1

SCHEMATIZATION OF CONTINUA RESULTING FROM CONCURRENCES OF VARYING COMBINATIONS OF TWO DIMENSIONS



The 180 selected items were randomized with the provision that no two items from the same scale should be adjacent. The mimeographed "Inventory of Choices" was administered to two groups of individuals: 92 high-school students in a required biology course, composed of four sections, roughly stratified by the school according to IQ and prior achievement; and 50 undergraduate university students in two sections of an education course. Responses were indicated on a Likert-type six-point scale: "strongly agree," "moderately agree," "slightly agree," "slightly disagree," "moderately disagree," and "strongly disagree."

Response categories were assigned the arbitrary weights, 5-0, with strongly agree to positively phrased items, or strongly disagree to negatively phrased items, being assigned the weight 5, etc. Where a respondent could not make a decision, he was instructed to cross out the number of that item. Such infrequent responses were assigned the intermediate weight, $2\frac{1}{2}$. Scores on each scale for each individual were obtained by summing these arbitrary weights.

Following Louis Guttman's "Cornell technique,"[4] twelve tables were laid out, one for each scale for each of the two samples separately. Each table was constructed with one column for each of six response categories of each item, and one row for each individual in the sample. Thus, since each scale had thirty items and each item contained six response categories, each table consisted of 180 columns and as many rows as individuals in the sample. The individuals were ordered, high to low, for each scale, according to their summated scores, and listed vertically down the left side of each table. The responses of each individual to each item were indicated by X's in the appropriate cells, the resulting tables containing all of the data gathered.

The array of responses to each item was double-dichotomized: vertically by combining response categories (e.g., 5-3 considered positive, $2\frac{1}{2}$-0 considered negative); horizontally at the point where the frequency of positive responses was greater above the line, and negative responses greater below the line. The responses in each of the four resulting quadrants were then tabulated in 2 × 2 contingency tables. The association between high scores (dependent on positive responses to a majority of items) and positive responses to each item, was tested by the chi-squared test corrected for continuity. Two values of chi-squared were thus obtained for each item—one value from each sample, providing replication for judging the reliability of the association in differing groups. Figure 2 illustrates these contingency tables.

All of the associations in the contingency tables were "positive," the values of chi-squared exceeding the .05 level of significance in the majority of cases. Examination of the array of responses to each item in the original tables, and their summarization in the contigency tables, points to the following observations: 1) while there was considerable "higgledy-piggledeliness"[5] in the response patterns, the central tendency of the responses to each item lay on a diagonal from upper left (high scores and high category values) to lower right (low scores and low category values); 2) the confirmation of the "prediction" of an individual's response to each item from his responses to all other[6] items confirms (a) the direction of the a priori weights originally assigned to the response categories of each item[7]; and (b) by the same token, the existence of a communality of content through all the items.

Twelve items were selected from the thirty of each scale on the basis of the following criteria:

1. The level of discrimination in the high-school group indicated by the value of chi-squared,
2. The reliability of the discrimination indicated by the replication with the university group,
3. The variation in the distribution of item marginal totals, indicating the "popularity" of positive responses, in order to provide several "cutting points" in the continua.[8]

The twelve selected items from each of the six scales were then again randomized to construct a revised "Inventory of Choices," composed of seventy-two items.

## Conditions for the "Measurement" of Attitudes

In a uni-dimensional scale, "... persons who answer a given question favorably all have higher ranks on the scale than persons who answer the same question unfavorably."[9] Alternatively, an individual who has a higher scale score than another, has given positive responses to all of the items which were positively responded to by the latter, plus one or more additional items. This requirement gives rise to Louis Guttman's stipulation that the responses to each item shall be a simple function of the persons' ranks:[10] that the responses to each item shall be reproducible from the scale score. If the items are arranged in order of popularity, a person with a scale score of 3 will have given positive responses to the three most popular items, etc. Thus, if we are to say that one individual is more "prudent" than another, we should mean that the former has given prudent responses to all the items responded prudently to by the latter, and more. If this condition be obtained, an "ordinal" or "rank order" scale would be constructed.[11]

With a perfectly cumulative scale, the frequencies in the upper-negative and lower-positive quadrants of the 2 × 2 contingency tables, such as are illustrated in Figure 2, would be zero. No

FIGURE 2

CONTINGENCY TABLES ILLUSTRATING THE ASSOCIATION OF POSITIVE RESPONSES WITH HIGH SCORES, AND NEGATIVE RESPONSES WITH LOW SCORES

Item 4

Sample A

| | Positive Responses | Negative Responses | Total |
|---|---|---|---|
| High Scores | 46 | 15 | 61 |
| Low Scores | 10 | 21 | 31 |
| Total | 56 | 36 | 92 |

$\chi^2 = 14.308$

Sample B

| | Positive Responses | Negative Responses | Total |
|---|---|---|---|
| High Scores | 21 | 4 | 25 |
| Low Scores | 8 | 17 | 25 |
| Total | 29 | 21 | 50 |

$\chi^2 = 11.822$

persons with scale scores above a certain point would give negative responses to an item, nor would any persons with scale scores below that point give positive responses to the item. Scores obtained with the use of such perfect scales would include no "errors." Designating obverse responses as "errors" does not mean that one could confront the respondent with his "error" and tell him that he was mistaken in his attitude—"You didn't really mean this." Each response must be assumed to define the respondent's attitude toward the content of the item. Hence these "errors" are departures of the data from the model of uni-dimensionality. They indicate multi-dimensionality in the content of the items. Confronted with the complexity of an item as a whole, the respondent's decision is based on the factors specific to each item—not necessarily on the factor common to all the items of a scale.

If two or three major dimensions exist in a scale then the interpretation of a given scale score is ambiguous since one respondent may have attained his score by positive responses to items of one dimension, while another respondent attained the same score by positive responses to items representing another dimension. For example, if income and education are combined as an index of socio-economic status, the meaning of a given rank is indeterminate since it may have been attained by either income or education, and these need not be interchangeable for the valid prediction of an external variable.[12]

On the other hand, if each item has specific factors, drawing idiosyncratic responses, but all of the items of a scale share a single common factor, then the rank order established by the group of items should rank individuals approximately in accord with the "ideal" ranking based on the common factor alone—specific factors more or less cancelling one another.

Since, in the formulation of the items for these scales, a deliberate effort was made to have each item represent a continuum in a different context, each item is obviously multi-dimensional. "Errors" are necessarily to be expected. But by examining the pattern of responses we should be able to make a preliminary judgment as to whether we have a single dominant factor and numerous specific factors, in which case the ordering of individuals would rest essentially on the dominant factor; or whether two or three major dimensions exist in the scale.

If a large group of individuals who have high scale scores—having given positive responses to a majority of items—give negative responses to one or two items, while a group of individuals with lower scale scores give positive responses to the same one or two items; then those items, having blocked groups of errors, represent a different dimension from the rest of the scale. On the other hand, if the frequency of negative responses to each item increases continuously as one descends the order of scale scores, then each item participates in a communality of content, and contributes to the discrimination of ranks in the same dimension.[13]

Were items representing a different dimension obversely related to the rest of the items in a linear fashion throughout the range of scale scores, so that the higher the scale score the greater the frequency of negative responses, then the obverse association in the condensed contingency tables, or the diminished value of chi-squared, would reveal this without a detailed examination of the pattern of responses through the range of scale scores. However, if such an item were positively related to the rest of the scale through the greater part of the range of scale scores, but obversely related in some limited range—as if, for example, those most strongly "prudent," along with those who are "immediate," as indicated by total scores, should give "immediate" responses to one item, while those who are moderately "prudent" should give the "prudent" responses—the condensation of these responses into a 2 × 2 table would obscure this curvilinear relationship, while a significant value of chi-squared might still be obtained.

We shall be concerned, in the preliminary analysis of the scales, therefore, not only with the association of high scale scores with positive responses to each item, to determine whether each item of the scale contributes to the discrimination of ranks in the same dimension; but also with the pattern of responses to determine whether this essential uni-dimensionality obtains throughout the range of scale scores.

Afterwards, if the scales prove to be essentially uni-dimensional, even though errors may be numerous, then the cancellation of specific factors may be enhanced, and an approximately cumulative scale constructed, by grouping equivalently popular items and defining positive responses to a majority of each group of items as a positive response to the compound item. By the use of this "H-technique" of Samuel Stouffer,[14] the dominant factor is brought out more clearly by making allowance for some idiosyncratic variation. This reduces the number of errors, thus enhancing the possibility of reproducing scores from ranks and vice versa, but it also reduces the number of possible ranks to which an individual may be assigned. In a sense, the technique helps the investigator use an instrument within its powers of resolution.

## Preliminary Analysis of the Revised Scales

The revised inventory of seventy-two items, consisting of the twelve selected items for each of the six scales, was administered to 325 high-school students in college-preparatory chemistry and physics classes.

The responses were again tabulated into six tables, using the "Cornell technique" as described above. The matrix of responses to each item was double-dichotomized and tabulated in 2 × 2 contingency tables. The association of high scale scores with positive responses was significant beyond the .001 level for each item. The values of chi-squared are reported in the last row of Tables I-VI for the items of each scale.

Each individual was re-scored, assigning the weights 1 and 0 to the new categories formed by the vertical dichotomization. This resulted in a possible range of scale scores from 12 (positive responses to all items) to 0 (negative responses to all items). The number of positive responses to each item given by the individuals with each scale score was then tabulated. These are converted into proportions and reported in the body of Tables I-VI. (The ultimate positive frequencies can be recovered from the product of the proportion and the frequency at each scale score: $p \cdot f_o$; and the negative frequencies from $(1 - p) \cdot f_o$.)

It can be seen from inspection that the proportions of positive responses to each item form an approximately continuous gradient, such that the lower the scale score, the smaller the proportion of positive responses, subject to sampling fluctuations.[15] Therefore we may judge that each item makes a significant contribution to the discrimination of ranks along the same dimension.

However, the frequency of errors, attributable to specific or random factors, and the presence of reversals in the gradient of proportions of positive responses in adjacent ranks (see footnote 15), indicates that the amalgamation of ranks by the combination of "equivalent" items would not be wasteful of any relevant discriminations.

## Re-Scoring by the "H-Technique"

In order to minimize the reflection in the scale scores of idiosyncratic responses to specific factors, "equivalent" items were combined to form the elements of compound items. The combination or grouping was done so that 1) the elements of each compound item should be as similar as possible to each other, and 2) the compound items should be as different as possible from one another.

The cumulative proportion of responses in each category of each item was computed and the twelve items of each scale were then ranked in order of increasing popularity of positive responses. Three compound items were formed from the items of each scale, the four least popular items forming the elements of compound item A, ...the four most popular items forming the elements of compound item C.

The respones to the elements of each compound item were then dichotomized so that approximately equal proportions of the responses to the elements of any one compound item should be positive. At the same time, about 25 percent of the respondents should give positive responses to three or four of the elements of compound item A, about 50 percent should give positive responses to three or four of the elements of compound item B, and about 75 percent should give positive responses to three or four of the elements of compound item C. In this way, the elements of each compound item are as equivalent as possible, and the resulting compound items are as different as possible from each other, "equivalence" and "difference" being defined by the population of respondents rather than by a priori judgment. This procedure is illustrated by the data from the prudent-theoretic scale in Table VII.

A positive response to each of the compound items is defined by a modal positive response to the elements of that item. In other words, positive responses to all four, or to three out of the four, elements of a compound item are scored as a positive response to the compound item; negative responses to two, three, or four of the four elements are scored as a negative response to the compound item.

The 325 individuals in the high-school group were then re-scored according to their modal responses to the compound items on each scale. The frequency of each possible pattern of response to the compound items of the six scales is shown in Table VIII.

## Analysis of the Modified Scales

It can be seen that the patterns of responses given by the majority of individuals were cumulative. If they gave a positive response to the least popular item, A, they also gave positive responses to the other two items; if they gave a positive response to item B, they also gave a positive response to C; and, vice versa, if they gave a negative response to C, they also gave a negative response to A and B, if they gave a negative response to B, they also gave a negative response to A. For these "scale-types" it is possible to say that an individual with a score of, for example, 2, made a positive response to the item responded to positively by an individual with a score of 1, and one more. Moreover, it is possible to determine, from the scale scores, which items an individual made positive responses to.

Inspection of the frequencies of non-cumulative response patterns shows that the greater the discrepancy of the pattern from the uni-dimensional pattern, the less frequently the pattern is observed. For example, of the 93 individuals who gave positive responses to only one item of the prudent-theoretic scale, 79 gave the positive response to item C, 12 to item B, and 2 to item A.[16]

The frequencies of positive and negative responses to each item by the individuals in each

## TABLE I

### PRUDENT-THEORETIC SCALE

PROPORTIONS OF POSITIVE RESPONSES GIVEN TO EACH ITEM BY THE INDIVIDUALS IN EACH RANK; TOGETHER WITH THE VALUES OF CHI-SQUARED, TESTING THE ASSOCIATION OF POSITIVE RESPONSES WITH UPPER RANKS

| | | Items | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rank | Frequency | 1 | 7 | 12 | 16 | 30 | 35 | 40 | 44 | 49 | 54 | 59 | 68 |
| 12 | 7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 11 | 17 | 1.000 | .941 | .765 | .941 | 1.000 | 1.000 | 1.000 | 1.000 | .941 | .941 | 1.000 | .471 |
| 10 | 27 | 1.000 | .852 | .815 | .630 | .852 | 1.000 | .963 | 1.000 | .852 | .815 | .852 | .370 |
| 9 | 31 | .968 | .677 | .613 | .484 | .903 | .935 | .742 | .871 | .935 | .903 | .710 | .258 |
| 8 | 34 | .912 | .588 | .529 | .618 | .824 | .735 | .735 | .912 | .735 | .618 | .706 | .088 |
| 7 | 30 | .867 | .533 | .367 | .533 | .633 | .833 | .600 | .800 | .700 | .500 | .600 | .033 |
| 6 | 24 | .583 | .292 | .417 | .458 | .583 | .792 | .542 | .792 | .625 | .417 | .417 | .083 |
| 5 | 29 | .655 | .345 | .379 | .414 | .552 | .517 | .379 | .621 | .448 | .276 | .345 | .069 |
| 4 | 33 | .212 | .242 | .242 | .455 | .485 | .667 | .152 | .545 | .273 | .242 | .455 | .030 |
| 3 | 27 | .259 | .148 | .296 | .333 | .259 | .519 | .111 | .296 | .259 | .111 | .333 | .074 |
| 2 | 28 | .214 | .000 | .143 | .464 | .321 | .357 | .071 | .071 | .143 | .036 | .107 | .071 |
| 1 | 26 | .154 | .038 | .154 | .231 | .231 | .000 | .000 | .077 | .038 | .000 | .038 | .038 |
| 0 | 12 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| Chi-squared | | 141.021 | 94.016 | 50.729 | 20.962 | 65.988 | 83.457 | 124.264 | 127.335 | 107.256 | 107.770 | 75.975 | 55.133 |

### TABLE II

### PRUDENT-IMMEDIATE SCALE

### PROPORTIONS OF POSITIVE RESPONSES GIVEN TO EACH ITEM BY THE INDIVIDUALS IN EACH RANK; TOGETHER WITH THE VALUES OF CHI-SQUARED, TESTING THE ASSOCIATION OF POSITIVE RESPONSES WITH UPPER RANKS

| | | Items | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rank | Frequency | 4 | 9 | 19 | 23 | 26 | 36 | 39 | 45 | 51 | 60 | 65 | 71 |
| 12 | 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 11 | 12 | 1.000 | .833 | 1.000 | 1.000 | .917 | .833 | .917 | 1.000 | .917 | .917 | .917 | .750 |
| 10 | 12 | 1.000 | .750 | .833 | .833 | .583 | .833 | .917 | 1.000 | .750 | 1.000 | .833 | .667 |
| 9 | 30 | .667 | .767 | .833 | .667 | .733 | .467 | .700 | .900 | .900 | .767 | .833 | .767 |
| 8 | 43 | .744 | .558 | .767 | .512 | .628 | .326 | .767 | .860 | .814 | .767 | .721 | .535 |
| 7 | 53 | .566 | .679 | .679 | .472 | .472 | .396 | .566 | .736 | .698 | .717 | .566 | .453 |
| 6 | 53 | .604 | .547 | .491 | .415 | .491 | .208 | .340 | .660 | .698 | .623 | .396 | .528 |
| 5 | 43 | .372 | .698 | .419 | .419 | .349 | .023 | .419 | .581 | .512 | .488 | .279 | .442 |
| 4 | 37 | .216 | .459 | .189 | .189 | .432 | .108 | .270 | .595 | .514 | .297 | .324 | .405 |
| 3 | 26 | .154 | .385 | .077 | .115 | .269 | .154 | .308 | .538 | .308 | .154 | .231 | .308 |
| 2 | 14 | .286 | .000 | .071 | .143 | .357 | .000 | .143 | .357 | .286 | .143 | .071 | .143 |
| 1 | 1 | .000 | 1.000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| Chi-squared | | 51.288 | 20.309 | 68.107 | 27.823 | 14.690 | 33.290 | 46.839 | 22.259 | 35.087 | 54.938 | 54.421 | 13.598 |

## TABLE III

## PRUDENT-AESTHETIC SCALE

PROPORTIONS OF POSITIVE RESPONSES GIVEN TO EACH ITEM BY THE INDIVIDUALS IN EACH RANK; TOGETHER WITH THE VALUES OF CHI-SQUARED, TESTING THE ASSOCIATION OF POSITIVE RESPONSES WITH UPPER RANKS

| Rank | Frequency | Items | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 3 | 8 | 20 | 27 | 29 | 34 | 42 | 48 | 52 | 57 | 64 | 70 |
| 12 | 6 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 11 | 18 | 1.000 | .889 | .944 | 1.000 | 1.000 | .833 | .833 | .889 | 1.000 | .944 | .833 | .833 |
| 10 | 25 | .840 | .960 | .720 | .880 | .720 | .880 | .800 | .760 | .880 | .920 | .680 | .960 |
| 9 | 33 | .939 | .697 | .818 | .879 | .727 | .485 | .727 | .515 | .848 | .818 | .636 | .909 |
| 8 | 56 | .893 | .714 | .661 | .714 | .607 | .375 | .732 | .571 | .750 | .679 | .607 | .696 |
| 7 | 44 | .818 | .773 | .545 | .591 | .500 | .227 | .568 | .477 | .773 | .591 | .523 | .614 |
| 6 | 51 | .627 | .667 | .627 | .471 | .333 | .157 | .373 | .353 | .706 | .608 | .549 | .529 |
| 5 | 30 | .567 | .667 | .400 | .467 | .400 | .200 | .367 | .333 | .467 | .400 | .300 | .433 |
| 4 | 34 | .412 | .647 | .265 | .294 | .265 | .118 | .324 | .176 | .471 | .235 | .294 | .500 |
| 3 | 18 | .278 | .278 | .333 | .111 | .111 | .000 | .167 | .222 | .333 | .333 | .167 | .667 |
| 2 | 6 | .500 | .000 | .333 | .000 | .000 | .000 | .000 | .167 | .000 | .167 | .500 | .333 |
| 1 | 4 | .000 | .250 | .250 | .250 | .000 | .000 | .000 | .000 | .000 | .000 | .250 | .000 |
| Chi-squared | | 52.292 | 15.958 | 23.910 | 56.180 | 47.323 | 71.789 | 53.072 | 33.751 | 48.600 | 38.776 | 20.628 | 26.117 |

## TABLE IV

### THEORETIC-IMMEDIATE SCALE

### PROPORTIONS OF POSITIVE RESPONSES GIVEN TO EACH ITEM BY THE INDIVIDUALS IN EACH RANK; TOGETHER WITH THE VALUES OF CHI-SQUARED, TESTING THE ASSOCIATION OF POSITIVE RESPONSES WITH UPPER RANKS

| | | Items | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rank | Frequency | 2 | 6 | 10 | 14 | 22 | 31 | 41 | 46 | 53 | 58 | 62 | 69 |
| 12 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 11 | 15 | .933 | .933 | .733 | .867 | .933 | 1.000 | .867 | 1.000 | .933 | .933 | .867 | 1.000 |
| 10 | 26 | .923 | .731 | .654 | .885 | .692 | .731 | .846 | .962 | .962 | .885 | .769 | .962 |
| 9 | 32 | .844 | .656 | .500 | .938 | .719 | .594 | .750 | .938 | .812 | .906 | .438 | .906 |
| 8 | 45 | .800 | .578 | .467 | .867 | .467 | .422 | .489 | .889 | .644 | .822 | .689 | .867 |
| 7 | 34 | .647 | .471 | .353 | .618 | .500 | .500 | .500 | .735 | .735 | .647 | .529 | .765 |
| 6 | 38 | .526 | .395 | .237 | .579 | .474 | .500 | .500 | .711 | .368 | .526 | .474 | .711 |
| 5 | 43 | .488 | .279 | .302 | .465 | .256 | .326 | .302 | .791 | .372 | .535 | .326 | .558 |
| 4 | 34 | .353 | .382 | .235 | .353 | .324 | .294 | .059 | .559 | .265 | .294 | .324 | .559 |
| 3 | 32 | .219 | .250 | .125 | .219 | .281 | .219 | .188 | .406 | .219 | .281 | .188 | .406 |
| 2 | 17 | .059 | .235 | .000 | .176 | .118 | .118 | .118 | .471 | .235 | .000 | .176 | .294 |
| 1 | 5 | .200 | .000 | .000 | .000 | .000 | .200 | .000 | .200 | .000 | .000 | .000 | .400 |
| 0 | 2 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| Chi-squared | | 60.310 | 36.254 | 35.760 | 68.500 | 31.974 | 30.778 | 61.968 | 48.116 | 77.126 | 68.091 | 34.671 | 38.656 |

## TABLE V

## THEORETIC-AESTHETIC SCALE

PROPORTIONS OF POSITIVE RESPONSES GIVEN TO EACH ITEM BY THE INDIVIDUALS IN EACH RANK; TOGETHER WITH THE VALUES OF CHI-SQUARED, TESTING THE ASSOCIATION OF POSITIVE RESPONSES WITH UPPER RANKS

| | | Items | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rank | Frequency | 13 | 17 | 21 | 25 | 28 | 33 | 38 | 47 | 56 | 61 | 66 | 72 |
| 12 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 11 | 22 | .955 | .909 | .864 | .955 | 1.000 | .955 | 1.000 | .955 | .682 | .818 | .955 | .955 |
| 10 | 42 | .857 | .881 | .881 | .881 | .881 | .762 | .905 | .881 | .643 | .738 | .810 | .881 |
| 9 | 48 | .813 | .646 | .812 | .708 | .875 | .667 | .917 | .792 | .500 | .729 | .708 | .833 |
| 8 | 39 | .744 | .590 | .821 | .641 | .795 | .436 | .821 | .744 | .564 | .641 | .564 | .641 |
| 7 | 51 | .706 | .471 | .608 | .608 | .686 | .490 | .725 | .667 | .431 | .549 | .412 | .647 |
| 6 | 46 | .652 | .370 | .565 | .543 | .609 | .413 | .652 | .630 | .326 | .478 | .326 | .435 |
| 5 | 29 | .552 | .483 | .310 | .724 | .483 | .207 | .414 | .414 | .207 | .276 | .517 | .414 |
| 4 | 28 | .571 | .179 | .250 | .571 | .214 | .214 | .464 | .321 | .464 | .286 | .179 | .286 |
| 3 | 11 | .091 | .182 | .182 | .455 | .364 | .182 | .182 | .364 | .273 | .273 | .091 | .364 |
| 2 | 6 | .167 | .000 | .000 | .833 | .000 | .333 | .167 | .167 | .000 | .000 | .167 | .167 |
| 1 | 1 | .000 | .000 | .000 | .000 | .000 | .000 | 1.000 | .000 | .000 | .000 | .000 | .000 |
| Chi-squared | | 26.322 | 45.068 | 55.728 | 11.164 | 53.661 | 45.373 | 51.607 | 36.767 | 18.632 | 34.687 | 50.716 | 50.801 |

## TABLE VI

## AESTHETIC-IMMEDIATE SCALE

PROPORTIONS OF POSITIVE RESPONSES GIVEN TO EACH ITEM BY THE INDIVIDUALS IN EACH RANK; TOGETHER WITH THE VALUES OF CHI-SQUARED, TESTING THE ASSOCIATION OF POSITIVE RESPONSES WITH UPPER RANKS

| Rank | Frequency | Items<br>5 | 11 | 15 | 18 | 24 | 32 | 37 | 43 | 50 | 55 | 63 | 67 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | .500 | 1.000 | .500 | 1.000 | 1.000 |
| 10 | 7 | .714 | .857 | .857 | 1.000 | 1.000 | 1.000 | .857 | .571 | .857 | .857 | .429 | 1.000 |
| 9 | 7 | .571 | .857 | .857 | 1.000 | .571 | .857 | .571 | .571 | .857 | .857 | .429 | .714 |
| 8 | 14 | .714 | .714 | .714 | .643 | .500 | .857 | .857 | .786 | .643 | .571 | .357 | .643 |
| 7 | 29 | .793 | .828 | .655 | .724 | .448 | .586 | .621 | .310 | .517 | .517 | .586 | .379 |
| 6 | 37 | .595 | .514 | .649 | .622 | .459 | .486 | .459 | .243 | .486 | .432 | .568 | .514 |
| 5 | 54 | .481 | .333 | .556 | .426 | .204 | .481 | .444 | .278 | .463 | .463 | .481 | .370 |
| 4 | 56 | .500 | .268 | .589 | .321 | .125 | .411 | .250 | .179 | .268 | .446 | .393 | .250 |
| 3 | 53 | .358 | .170 | .491 | .189 | .038 | .321 | .132 | .245 | .264 | .321 | .245 | .208 |
| 2 | 45 | .178 | .067 | .267 | .200 | .067 | .178 | .111 | .133 | .289 | .244 | .200 | .067 |
| 1 | 20 | .050 | .000 | .050 | .000 | .100 | .250 | .200 | .050 | .000 | .150 | .050 | .200 |
| 0 | 1 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| Chi-squared | | 26.852 | 73.092 | 26.601 | 57.057 | 62.276 | 25.344 | 45.510 | 13.114 | 29.640 | 11.700 | 21.528 | 31.200 |

## TABLE VII

## CUMULATIVE PROPORTIONS OF RESPONSES IN EACH CATEGORY OF EACH ITEM OF THE PRUDENT-THEORETIC SCALE

| | Contrived Items* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | | | | B | | | | C | | | |
| Category ''weight'' | 68 | 12 | 49 | 7 | 54 | 44 | 40 | 16 | 59 | 30 | 1 | 35 |
| 5 | .05 | .07 | .11 | .12 | .07 | .18 | .18 | .17 | .17 | .32 | .21 | .27 |
| 4 | .14 | .16 | .23 | .27 | .25 | .27 | .31 | .34 | .33 | .43 | .45 | .44 |
| 3 | .25 | .27 | .37 | .40 | .42 | .44 | .46 | .49 | .49 | .58 | .60 | .65 |
| 2½ | .31 | .28 | .38 | .41 | .43 | .47 | .50 | .50 | .53 | .60 | .61 | .67 |
| 2 | .50 | .42 | .52 | .53 | .61 | .62 | .61 | .64 | .73 | .72 | .72 | .77 |
| 1 | .75 | .62 | .73 | .75 | .83 | .74 | .80 | .82 | .85 | .83 | .88 | .87 |
| 0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

*Items are arranged in order of ''popularity'' (as measured by the cumulative proportion of responses in categories five through three), and grouped to form ''contrived items''. Horizontal lines indicate the dichotomization of the elements into positive and negative categories.

TABLE VIII

FREQUENCY OF RESPONSES OF EACH PATTERN TO THE ITEMS OF EACH SCALE

| Score | Response Pattern ABC | Frequency in Each Scale | | | | | |
|---|---|---|---|---|---|---|---|
| | | P-T | P-I | P-A | T-I | T-A | A-I |
| 3 | + + + | 61 | 42 | 40 | 48 | 42 | 49 |
| 2 | - + + | 88 | 103 | 62 | 87 | 104 | 78 |
| | * + - + | 11 | 15 | 19 | 15 | 20 | 15 |
| | * + + - | 0 | 4 | 3 | 1 | 3 | 5 |
| 1 | - - + | 79 | 107 | 119 | 73 | 83 | 107 |
| | * - + - | 12 | 17 | 15 | 12 | 18 | 12 |
| | * + - - | 2 | 2 | 4 | 5 | 6 | 7 |
| 0 | - - - | 72 | 35 | 63 | 84 | 49 | 52 |

*Non-cumulative response patterns, classed as errors.

rank are recorded in the condensed scalograms of Tables IX-XIV. For item A of each scale positive responses by individuals having scores of 2-0 are "errors"; for item B, negative responses by individuals having a score of 2 and positive responses by individuals having a score of 1, are "errors," etc.

A coefficient of reproducibility is the proportion of responses which fit the cumulative patterns. A coefficient of reproducibility of .92 means that from an individual's scale score we can reproduce his responses to each item with 92 percent accuracy (by predicting the cumulative pattern of responses). The coefficients of reproducibility for each scale are reported in the "total" column of the last rows of Tables IX-XIV. The figures under each item apply to the items separately. These coefficients are seen to range from .90-.95, which values are within the bounds established by Guttman for adequate reproducibility.[17]

Thus, within the limits of error described by the coefficient of reproducibility, the scale scores validly order the individuals along the continua defined by the content of the items. It is possible, with 90-95 percent accuracy, to say, for example, that a more "prudent" individual has made prudent responses to all items responded to prudently by a less prudent individual, and more.

The coefficient of reproducibility describes the confidence with which we may assume that a given scale score represents a particular cumulative pattern of responses. But it does not describe the confidence with which we may reject the possibility that the cumulative pattern of responses might have been obtained by chance were the items actually independent rather than cumulatively interdependent.

For example, with three independent items, if 3/4 answer the most popular item, C, positively, 1/2 answer the next most popular item, B, positively, and 1/4 answer the least popular item, A, positively; then 1/2 of 3/4 = 3/8 would answer both B and C positively by chance, if the items had no relationship, and 1/4 of 3/8 = 3/32 would answer all three positively by chance, etc. Thus, the expectation for each response pattern for independent items can be computed from the product of the proportions giving each response.

If the items are strictly cumulative, however, then all those who give positive responses to A will also give positive responses to B and C. Thus 1/4 or 8/32, rather than 3/32, will give positive responses to all three items; none will give non-cumulative response patterns, etc.

The expected proportions obtained from these hypothetical distributions under the two assumptions of 1) independence of items, the null hypothesis, $H_0$, and 2) cumulative interdependence of items, $H_c$, are tabulated in Figure 3. Comparing the proportions under $H_0$ and $H_c$ we can predict the direction, or sign, of differences to be anticipated between the expected frequencies under $H_0$ and actually observed frequencies, if the direction of departure from independence is toward cumula-

## TABLE IX

## PRUDENT-THEORETIC SCALE

FREQUENCY OF POSITIVE AND NEGATIVE RESPONSES TO EACH ITEM OF THE PRUDENT-THEORETIC SCALE BY THE INDIVIDUALS OF EACH SCALE SCORE

| | | Items | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | A | | B | | C | | |
| Score | Frequency | Positive | Negative | Positive | Negative | Positive | Negative | Total |
| 3 | 61 | 61 | 0 | 61 | 0 | 61 | 0 | 183 |
| 2 | 99 | 11 | 88 | 88 | 11 | 99 | 0 | 297 |
| 1 | 93 | 2 | 91 | 12 | 81 | 79 | 14 | 279 |
| 0 | 72 | 0 | 72 | 0 | 72 | 0 | 72 | 216 |
| Total | 325 | 74 | 251 | 161 | 164 | 239 | 86 | 975 |
| Errors | | 13 | 0 | 12 | 11 | 0 | 14 | 50 |
| C of R | | .96 | | .93 | | .96 | | .95 |

tive interdependence. These signs are listed in the last column of Figure 3 for the proportions used here as an example.

In Tables XV-XX the chi-squared test has been applied to the goodness of fit of the obtained frequencies to the expected frequencies under the assumption of independence, $H_0$, for each of the six scales. In each case the probability of obtaining the observed distribution, were the items actually independent, was less than .001. Therefore, we can reject the possibility that the items are independent with the confidence that we should be wrong in making such decisions fewer than one time in a thousand.

Moreover, the direction of the interdependence is substantially in the direction predicted under the assumption of cumulative interdependence. The contributions to the value of chi-squared of differences which are not in the predicted direction have been segregated into the last column of each of the tables. They can be seen to be negligible—in no case affecting the level of significance of the value of chi-squared.

### Summary

It was the purpose of this research project to construct an instrument whereby individuals might be validly ordered along six theoretical continua. It was determined that "order" should be assumed to exist only if individuals' decision behavior evinced a particular cumulative structure, so that a person at any given position on the continuum would make the "favorable" decisions made by all persons "lower" on the continuum, plus one or more additional "favorable" decisions.

Because of the complexity of considerations involved in the process of making decisions in the variety of contexts offered as "items," there was considerable "higgledy-piggledidiness" in the patterns of the responses; the many discriminations being based on differences in a large number of factors, rather than cumulative differences in a uni-dimensional factor.

By a process of repeated consolidation of responses: first, by dichotomizing the original six responses into positive and negative responses, and then by combining items, and dichotomizing the responses to the elements into positive and negative responses to the compound items, it was found possible to eliminate most of the discriminations based upon idiosyncratic factors.

The remaining discriminations coarsely ranked the individuals in a cumulative order, so that one might say that an individual with a given scale

FIGURE 3

A PRIORI EXPECTED PROPORTIONS UNDER THE ALTERNATIVE HYPOTHESES OF INDEPENDENT ATTRIBUTES AND OF CUMULATIVE ATTRIBUTES

| Response Pattern | Expected Proportion $H_0$ | Expected Proportion $H_c$ | Predicted Sign of Difference |
|---|---|---|---|
| + + + | $A \cdot B \cdot C = \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{32}$ | $A = \frac{1}{4} = \frac{8}{32}$ | + |
| - + + | $\alpha \cdot B \cdot C = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{3}{4} = \frac{9}{32}$ | $B - A = \frac{1}{4} = \frac{8}{32}$ | - |
| + - + | $A \cdot \beta \cdot C = \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{32}$ | 0 | - |
| + + - | $A \cdot B - \gamma = \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{32}$ | 0 | - |
| - - + | $\alpha \cdot \beta \cdot C = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{3}{4} = \frac{9}{32}$ | $\beta - \gamma = \frac{1}{4} = \frac{8}{32}$ | - |
| - + - | $\alpha \cdot B \cdot \gamma = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{32}$ | 0 | - |
| + - - | $A \cdot \beta \cdot \gamma = \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{32}$ | 0 | - |
| - - - | $\alpha \cdot \beta \cdot \gamma = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{32}$ | $\gamma = \frac{1}{4} = \frac{8}{32}$ | + |

Note: A indicates the proportion of positive responses to item A; B, the positive responses to B; C, the positive responses to C; $\alpha$, the negative responses to A; $\beta$, the negative responses to B; $\gamma$, the negative responses to C.

## TABLE X

## PRUDENT-IMMEDIATE SCALE

FREQUENCY OF POSITIVE AND NEGATIVE RESPONSES TO EACH ITEM OF THE PRUDENT-IMMEDIATE SCALE BY THE INDIVIDUALS OF EACH SCALE SCORE

| | | Items | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | A | | B | | C | | |
| Score | Frequency | Positive | Negative | Positive | Negative | Positive | Negative | Total |
| 3 | 42 | 42 | 0 | 42 | 0 | 42 | 0 | 126 |
| 2 | 122 | 19 | 103 | 107 | 15 | 118 | 4 | 366 |
| 1 | 126 | 2 | 124 | 17 | 109 | 107 | 19 | 378 |
| 0 | 35 | 0 | 35 | 0 | 35 | 0 | 35 | 105 |
| Total | 325 | 63 | 262 | 166 | 159 | 267 | 58 | 975 |
| Errors | | 21 | 0 | 17 | 15 | 0 | 23 | 76 |
| C of R | | .94 | | .90 | | .93 | | .92 |

## TABLE XI

## PRUDENT-AESTHETIC SCALE

FREQUENCY OF POSITIVE AND NEGATIVE RESPONSES TO EACH ITEM OF THE PRUDENT-AESTHETIC SCALE BY THE INDIVIDUALS OF EACH SCALE SCORE

| | | Items | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | A | | B | | C | | |
| Score | Frequency | Positive | Negative | Positive | Negative | Positive | Negative | Total |
| 3 | 40 | 40 | 0 | 40 | 0 | 40 | 0 | 120 |
| 2 | 84 | 22 | 62 | 65 | 19 | 81 | 3 | 252 |
| 1 | 138 | 4 | 134 | 15 | 123 | 119 | 19 | 414 |
| 0 | 63 | 0 | 63 | 0 | 63 | 0 | 63 | 189 |
| Total | 325 | 66 | 259 | 120 | 205 | 240 | 85 | 975 |
| Errors | | 26 | 0 | 15 | 19 | 0 | 22 | 82 |
| C of R | | .92 | | .90 | | .93 | | .92 |

## TABLE XII

## THEORETIC-IMMEDIATE SCALE

FREQUENCY OF POSITIVE AND NEGATIVE RESPONSES TO EACH ITEM OF THE THEORETIC-IMMEDIATE SCALE BY THE INDIVIDUALS OF EACH SCALE SCORE

| | | Items | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | A | | B | | C | | |
| Score | Frequency | Positive | Negative | Positive | Negative | Positive | Negative | Total |
| 3 | 48 | 48 | 0 | 48 | 0 | 48 | 0 | 144 |
| 2 | 103 | 16 | 87 | 88 | 15 | 102 | 1 | 309 |
| 1 | 90 | 5 | 85 | 12 | 78 | 73 | 17 | 270 |
| 0 | 84 | 0 | 84 | 0 | 84 | 0 | 84 | 252 |
| Total | 325 | 69 | 256 | 148 | 177 | 223 | 102 | 975 |
| Errors | | 21 | 0 | 12 | 15 | 0 | 18 | 66 |
| C of R | | .94 | | .92 | | .94 | | .93 |

## TABLE XIII

## THEORETIC-AESTHETIC SCALE

FREQUENCY OF POSITIVE AND NEGATIVE RESPONSES TO EACH ITEM OF THE THEORETIC-AESTHETIC SCALE BY THE INDIVIDUALS OF EACH SCALE SCORE

| | | Items | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | A | | B | | C | | |
| Score | Frequency | Positive | Negative | Positive | Negative | Positive | Negative | Total |
| 3 | 42 | 42 | 0 | 42 | 0 | 42 | 0 | 126 |
| 2 | 127 | 23 | 104 | 107 | 20 | 124 | 3 | 381 |
| 1 | 107 | 6 | 101 | 18 | 89 | 83 | 24 | 321 |
| 0 | 49 | 0 | 49 | 0 | 49 | 0 | 49 | 147 |
| Total | 325 | 71 | 254 | 167 | 158 | 249 | 76 | 975 |
| Errors | | 29 | 0 | 18 | 20 | 0 | 27 | 94 |
| C of R | | .91 | | .88 | | .92 | | .90 |

## TABLE XIV

### AESTHETIC-IMMEDIATE SCALE

FREQUENCY OF POSITIVE AND NEGATIVE RESPONSES TO EACH ITEM OF THE AESTHETIC-IMMEDIATE SCALE BY THE INDIVIDUALS OF EACH SCALE SCORE

| | | Items | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | A | | B | | C | | |
| Score | Frequency | Positive | Negative | Positive | Negative | Positive | Negative | Total |
| 3 | 49 | 49 | 0 | 49 | 0 | 49 | 0 | 147 |
| 2 | 98 | 20 | 78 | 83 | 15 | 93 | 5 | 294 |
| 1 | 126 | 7 | 119 | 12 | 114 | 107 | 19 | 378 |
| 0 | 52 | 0 | 52 | 0 | 52 | 0 | 52 | 156 |
| Total | 325 | 76 | 249 | 144 | 181 | 249 | 76 | 975 |
| Errors | | 27 | 0 | 12 | 15 | 0 | 24 | 78 |
| C of R | | .92 | | .92 | | .93 | | .92 |

BUREAU OF EDNL. & PSYL. RESEARCH LIBRARY
9-306
3.9.70

### TABLE XV

### PRUDENT-THEORETIC SCALE

### CHI-SQUARED TEST OF THE SIGNIFICANCE OF DISCREPANCIES BETWEEN OBTAINED FREQUENCIES AND EXPECTED FREQUENCIES UNDER THE ASSUMPTION OF INDEPENDENCE

| Response Pattern | $f_o$ Observed Frequency | $f_e(H0)$ Expected Frequency | Predicted Sign of Difference | Obtained Difference | $\frac{(f_o-f_e)^2}{f_e}$ | Contribution of Unpredicted Differences |
|---|---|---|---|---|---|---|
| + + + | 61 | 26.95806 | + | +34.04194 | 42.98728 | |
| - + + | 88 | 91.43885 | - | - 3.43885 | 0.12932 | |
| * + - + | 11 | 27.46039 | - | -16.46039 | 9.86673 | |
| * + + - | 0 | 9.70039 | - | - 9.70039 | 9.70039 | |
| - - + | 79 | 93.14268 | - | -14.14268 | 2.14740 | |
| * - + - | 12 | 32.90268 | - | -20.90268 | 13.27922 | |
| * + - - | 2 | 9.88114 | - | - 7.88114 | 6.28595 | |
| - - - | 72 | 33.51577 | + | +38.48423 | 44.18922 | |
| Total | 325 | 324.99996 | | 0.00004 | 128.58551 | 0 |
| Degrees of Freedom[17] | | | | | 4 | |
| Probability less than | | | | | .001 | |

*Non-cumulative response patterns.

## TABLE XVI

## PRUDENT-IMMEDIATE SCALE

## CHI-SQUARED TEST OF THE SIGNIFICANCE OF DISCREPANCIES BETWEEN OBTAINED FREQUENCIES AND EXPECTED FREQUENCIES UNDER THE ASSUMPTION OF INDEPENDENCE

| Response Pattern | $f_o$ Observed Frequency | $f_e(H_0)$ Expected Frequency | Predicted Sign of Difference | Obtained Difference | $\frac{(f_o-f_e)^2}{f_e}$ | Contribution of Unpredicted Differences |
|---|---|---|---|---|---|---|
| + + + | 42 | 26.43584 | + | +15.56416 | 9.16343 | |
| - + + | 103 | 109.93954 | - | - 6.93954 | 0.43803 | |
| * + - + | 15 | 25.32107 | - | -10.32107 | 4.20695 | |
| * + + - | 4 | 5.74261 | - | - 1.74261 | 0.52879 | |
| - - + | 107 | 105.30353 | - | + 1.69647 | 0.02733 | 0.02733 |
| * - + - | 17 | 23.88199 | - | - 6.88199 | 1.98315 | |
| * + - - | 2 | 5.50045 | - | - 3.50045 | 2.22766 | |
| - - - | 35 | 22.87492 | + | +12.12508 | 6.42701 | |
| Total | 325 | 324.99995 | | 0.00005 | 25.00235 | 0.02733 |
| Degrees of Freedom | | | | | 4 | |
| Probability less than | | | | | .001 | |

*Non-cumulative response patterns.

## TABLE XVII

## PRUDENT-AESTHETIC SCALE

## CHI-SQUARED TEST OF THE SIGNIFICANCE OF DISCREPANCIES BETWEEN OBTAINED FREQUENCIES AND EXPECTED FREQUENCIES UNDER THE ASSUMPTION OF INDEPENDENCE

| Response Pattern | $f_o$ Observed Frequency | $f_e(H_0)$ Expected Frequency | Predicted Sign of Difference | Obtained Difference | $\frac{(f_o-f_e)^2}{f_e}$ | Contribution of Unpredicted Differences |
|---|---|---|---|---|---|---|
| + + + | 40 | 17.99573 | + | +22.00427 | 26.90571 | |
| - + + | 62 | 70.61964 | - | - 8.61964 | 1.05208 | |
| * + - + | 19 | 30.74272 | - | -11.74272 | 4.48533 | |
| * + + - | 3 | 6.37349 | - | - 3.37349 | 1.78558 | |
| - - + | 119 | 120.64189 | - | - 1.64189 | 0.02234 | |
| * - + - | 15 | 25.01112 | - | -10.01112 | 4.00711 | |
| * + - - | 4 | 10.88804 | - | - 6.88804 | 4.35754 | |
| - - - | 63 | 42.72733 | + | +20.27267 | 9.61869 | |
| Total | 325 | 324.99996 | | 0.00004 | 52.23438 | 0 |
| Degrees of Freedom | | | | | 4 | |
| Probability less than | | | | | .001 | |

*Non-cumulative response patterns.

## TABLE XVIII

## THEORETIC-IMMEDIATE SCALE

## CHI-SQUARED TEST OF THE SIGNIFICANCE OF DISCREPANCIES BETWEEN OBTAINED FREQUENCIES AND EXPECTED FREQUENCIES UNDER THE ASSUMPTION OF INDEPENDENCE

| Response Pattern | $f_o$ Observed Frequency | $f_e(H_0)$ Expected Frequency | Predicted Sign of Difference | Obtained Difference | $\frac{(f_o-f_e)^2}{f_e}$ | Contribution of Unpredicted Differences |
|---|---|---|---|---|---|---|
| + + + | 48 | 21.56000 | + | +26.44000 | 32.42456 | |
| - + + | 87 | 79.99075 | - | + 7.00925 | 0.61419 | 0.61419 |
| * + - + | 15 | 25.78460 | - | -10.78460 | 4.51073 | |
| * + + - | 1 | 9.86152 | - | - 8.86152 | 7.96292 | |
| - - + | 73 | 95.66462 | - | -22.66462 | 5.36964 | |
| * - + - | 12 | 36.58770 | - | -24.58770 | 16.52344 | |
| * + - - | 5 | 11.79385 | - | - 6.79385 | 3.91359 | |
| - - - | 84 | 43.75691 | + | +40.24309 | 37.01144 | |
| Total | 325 | 324.99995 | | 0.00005 | 108.33051 | 0.61419 |
| Degrees of Freedom | | | | | 4 | |
| Probability less than | | | | | .001 | |

*Non-cumulative response patterns.

TABLE XIX

THEORETIC-AESTHETIC SCALE

CHI-SQUARED TEST OF THE SIGNIFICANCE OF DISCREPANCIES BETWEEN OBTAINED FREQUENCIES AND EXPECTED FREQUENCIES UNDER THE ASSUMPTION OF INDEPENDENCE

| Response Pattern | $f_o$ Observed Frequency | $f_e(H_0)$ Expected Frequency | Predicted Sign of Difference | Obtained Difference | $\frac{(f_o-f_e)^2}{f_e}$ | Contribution of Unpredicted Differences |
|---|---|---|---|---|---|---|
| + + + | 42 | 27.95164 | + | +14.04836 | 7.06063 | |
| - + + | 104 | 99.99604 | - | + 4.00396 | 0.16032 | 0.16032 |
| * + - + | 20 | 26.44527 | - | - 6.44527 | 1.57084 | |
| * + + - | 3 | 8.53142 | - | - 5.53142 | 3.58634 | |
| - - + | 83 | 94.60703 | - | -11.60703 | 1.42402 | |
| * - + - | 18 | 30.52088 | - | -12.52088 | 5.13656 | |
| * + - - | 6 | 8.07164 | - | - 2.07164 | 0.53170 | |
| - - - | 49 | 28.87604 | + | +20.12396 | 14.02456 | |
| Total | 325 | 324.99996 | | 0.00004 | 33.49497 | 0.16032 |
| Degrees of Freedom | | | | | 4 | |
| Probability less than | | | | | .001 | |

*Non-cumulative response patterns.

## TABLE XX

## AESTHETIC-IMMEDIATE SCALE

## CHI-SQUARED TEST OF THE SIGNIFICANCE OF DISCREPANCIES BETWEEN OBTAINED FREQUENCIES AND EXPECTED FREQUENCIES UNDER THE ASSUMPTION OF INDEPENDENCE

| Response Pattern | $f_o$ Observed Frequency | $f_e(H_0)$ Expected Frequency | Predicted Sign of Difference | Obtained Difference | $\frac{(f_o-f_e)^2}{f_e}$ | Contribution of Unpredicted Differences |
|---|---|---|---|---|---|---|
| + + + | 49 | 25.79934 | + | +23.20066 | 20.86373 | |
| - + + | 78 | 84.52680 | - | - 6.52680 | 0.50397 | |
| * + - + | 15 | 32.42834 | - | -17.42834 | 9.36671 | |
| * + + - | 5 | 7.87449 | - | - 2.87449 | 1.04929 | |
| - - + | 107 | 106.24550 | - | + 0.75450 | 0.00535 | 0.00535 |
| * - + - | 12 | 25.79934 | - | -13.79934 | 7.38087 | |
| * + - - | 7 | 9.89780 | - | - 2.89780 | 0.84839 | |
| - - - | 52 | 32.42834 | + | +19.57166 | 11.81219 | |
| Total | 325 | 324.99995 | | 0.00005 | 51.83050 | 0.00535 |
| Degrees of Freedom | | | | | 4 | |
| Probability less than | | | | | .001 | |

*Non-cumulative response patterns

score made all of the "favorable" responses of individuals with lower scale scores, plus one or more additional "favorable" responses, with 90-95 percent accuracy. It was possible to reject, with a high degree of confidence ($p < .001$), the possibility that these cumulative patterns of response could have arisen with such frequency by chance were the items independent.

The instrument has validity, based upon its internal cumulative structure, for the rank ordering of individuals along the six continua, which should be of value in relating these variables with any external variables.

## FOOTNOTES

1. T. Bentley Edwards and Alan B. Wilson. "Attitudes Toward the Study of School Subjects," *Educational Theory*, VIII (October 1958), pp. 275-84.

2. For descriptive expansion of these vectors and their "ideal types" see Edwards and Wilson, *ibid.*

3. E.g., Allen L. Edwards. *Techniques of Attitude Scale Construction* (New York: Appleton-Century-Crofts, Inc., 1957), pp. 13-14; Katherine Gordon Capt, "The Questionnaire and Other Reporting Forms as Aids in Field Exploration," *Scientific Social Surveys and Research*, ed., by Pauline V. Young (Englewood Cliffs: Prentice-Hall, 1956), pp. 184-91.

4. Louis Guttman. "The Cornell Technique for Scale and Intensity Analysis," *Educational and Psychological Measurements*, VII (Summer 1947), pp. 247-79.

5. D. A. Walker. "Answer-Pattern and Score-Scatter in Tests and Examinations," *British Journal of Psychology*, XXII (July 1931), pp. 73-86.

6. Ideally each "prediction" should be based on the individual's total score excluding the contribution of the item being predicted. The lack of independence resulting from the inclusion of the "predicted" score in the total score does not substantially alter its significance where the number of items is large.

7. Cf. Shirley A. Star. "The Screening of Psychoneurotics in the Army: Technical Development of Tests," *Studies in Social Psychology in World War II, Vol. IV, Measurement and Prediction*, ed., by *Samuel A. Stouffer* (Princeton: Princeton University Press, 1950), p. 494: "... in no case did the scale or quasi-scale scoring of an item prove to be the reverse of the empirically found differences on that item. This outcome indicated, first, that our a priori judgment had been correct and, second, that the scale scoring, even on a dichotomous basis, preserved the predictive efficiency of the individual items...."

8. Since the phi-coefficient is directly proportional to chi-squared when N is constant, this basis of selection is analogous to that recommended by Allen L. Edwards and F. P. Kilpatrick, "A Technique for the Construction of Attitude Scales," *Journal of Applied Psychology*, XXXII (1948), 374-84, except that observed marginal totals rather than Thurstone scale values were used to obtain a range of items. Actually this last criterion was applied loosely, and, as a result, many "middle" items were selected, discriminating at the same level. This condition was corrected by the redefinition of positive responses in the subsequent improvement of the scales, infra.

9. Samuel A. Stouffer. "An Overview of the Contributions to Scaling and Scale Theory," *Measurement and Prediction*, op. cit., p. 9. Cf. Jane Loevinger. "A Systematic Approach to the Construction and Evaluation of Tests of Ability," *Psychological Monographs*, LXI (1947), No. 285; and "The Technic of Homogeneous Tests Compared with Some Aspects of 'Scale Analysis' and Factor Analysis," *Psychological Bulletin*, XLV (1948), pp. 507-29; G. Murphy, L. B. Murphy, and T. M. Newcomb, *Experimental Social Psychology* (New York: Harper and Bros., 1937), p. 897.

10. Louis Guttman. "A Basis for Scaling Qualitative Data," *American Sociological Review*, IX (April 1944), pp. 139-50; "The Basis for Scalogram Analysis," *Measurement and Prediction*, *op. cit.*, pp. 60-90.

11. John Dewey. *Logic: The Theory of Inquiry* (New York: Henry Holt and Co., 1938), p. 209: "A *measured* collection is identical with the *kind of collection* just said to have the property of totality in contrast with a *merely* numbered aggregate.... Measured collections involve (1) limits from which to which; (2) something specified as a unit for counting; and (3) progressive accumulation of these units until the limit *ad quem* is reached. The word *accumulation* as here used involves something different from the *aggregation* found in the merely

numerical set....

"The cumulative aspect in genuine collective propositions signifies that such propositions depend upon some principle of arrangement or order which is derived from the involved means-consequences relation...."

Cf. also with Morris R. Cohen and Ernest Nagel, An Introduction to Logic and the Scientific Method (New York: Harcourt, Brace and Co., 1934), especially Chapter XV, Section 5, "The Formal Conditions for Measurement," pp. 297-98.

12. Qualitatively diverse elements may be operationally equated and give rise to actuarial indexes which prove valid for a specific problem of external prediction. However, they are not theoretically pregnant since the grounds for their validity is ambiguous, i.e., they are not internally valid in the sense of "measuring" what they purport to measure. Interpretation leads to fallacious "inverse operationism." Cf. Clyde H. Coombs, "Theory and Methods of Social Measurement," Research Methods in the Behavioral Sciences, ed. by Leon Festinger and Daniel Katz (New York: Dryden Press, 1953), pp. 471-535.

13. Cf. Louis Guttman. "The Utility of Scalogram Analysis," Measurement and Prediction, op.cit., p. 159: "... This gradient pattern of errors indicates that, while there is not a single factor operating as in the case of a scale, nevertheless there is a single dominant factor and indefinitely many small random factors, so that the prediction of any external variable must rest essentially on the dominant factor..." See also ibid., pp. 160-63, 207, 208, 458, 494, and 547.

14. Samuel A. Stouffer, and others. "A Technique for Improving Cumulative Scales," Sociological Studies in Scale Analysis, ed. by Matilda White Riley and others (New Brunswick: Rutgers University Press, 1954), pp. 372-89.

15. In each case where a higher proportion of the individuals in the lower of two adjacent ranks gave positive responses to an item, the frequency of positive and negative responses in the two ranks was tabulated into a 2 × 2 contingency table. The null hypothesis $H_0$, that the difference in proportions of positive and negative responses between the two groups, in this reversed direction, could have occurred by chance were there no relevant differences between the ranks, was tested by: 1) the chi-squared test where the total number of individuals was greater than 40; or between 20 and 40, provided that the expected frequency in each cell was greater than 5; or 2) the Fisher-Yates test of exact probability where the number was less than 20; or between 20 and 40, where the expected frequency in some cell was less than 5.

In all but two of the cases it was impossible to reject $H_0$ at the .05 level. Hence, the null hypothesis is tenable for those cases. The two exceptions are shown in Figure 4, next page.

Reversals of this order, which could have occurred approximately 2.5 times in a hundred by chance, occurred twice in the 860 pairs of adjacent ranks of the six scales. However, the presence of these reversals confirms the fact, which our interpretation of errors provides further grounds for affirming, that the fineness of discrimination provided by the separate ranks are not valid, and may not be relevant to the common factors in the continua.

16. The "image" of each non-cumulative pattern is a cumulative pattern—in most instances of the next lower scale score. See Louis Guttman, "Image Theory of the Structure of Quantitative Variates," Psychometrika, XVIII (December 1953), pp. 277-79; and "The Israel Alpha Technique for Scale Analysis," Sociological Studies in Scale Analysis, op.cit., pp. 410-15.

17. The fact that assignment of frequencies to each of the eight response patterns can be made with only four degrees of freedom may not be immediately apparent. A three-dimensional contingency table, 2 × 2 × 2, provides the eight cells needed to house all possible combinations of positive and negative responses to each of the three items. This is illustrated in Figure 5.

Given the marginal totals of the frequency of positive and negative responses to each of the three items, then the joint-frequencies in only four cells may be assigned freely (provided the sum of the frequencies assigned to cells in the same plane do not exceed the marginal total of that plane), the frequencies in the remaining four cells being then determined by the marginal totals.

For example, the front face of the cube holds the four cells, a, b, c, and d, which house the frequencies of the positive responses to A. When three of these cells have been arbitrarily filled, using three degrees of freedom, the frequency in the

FIGURE 4

CONTINGENCY TABLES OF THE FREQUENCIES OF POSITIVE AND NEGATIVE RESPONSES GIVEN BY INDIVIDUALS IN ADJACENT RANKS WHERE A SIGNIFICANTLY GREATER ($p < .05$) PROPORTION OF THE INDIVIDUALS IN THE LOWER RANK GAVE POSITIVE RESPONSES

Item 56

| | Positive Responses | Negative Responses | Total |
|---|---|---|---|
| Rank 5 | 6 | 23 | 29 |
| Rank 4 | 13 | 15 | 28 |
| Total | 19 | 38 | 57 |

$\chi^2 = 3.167$

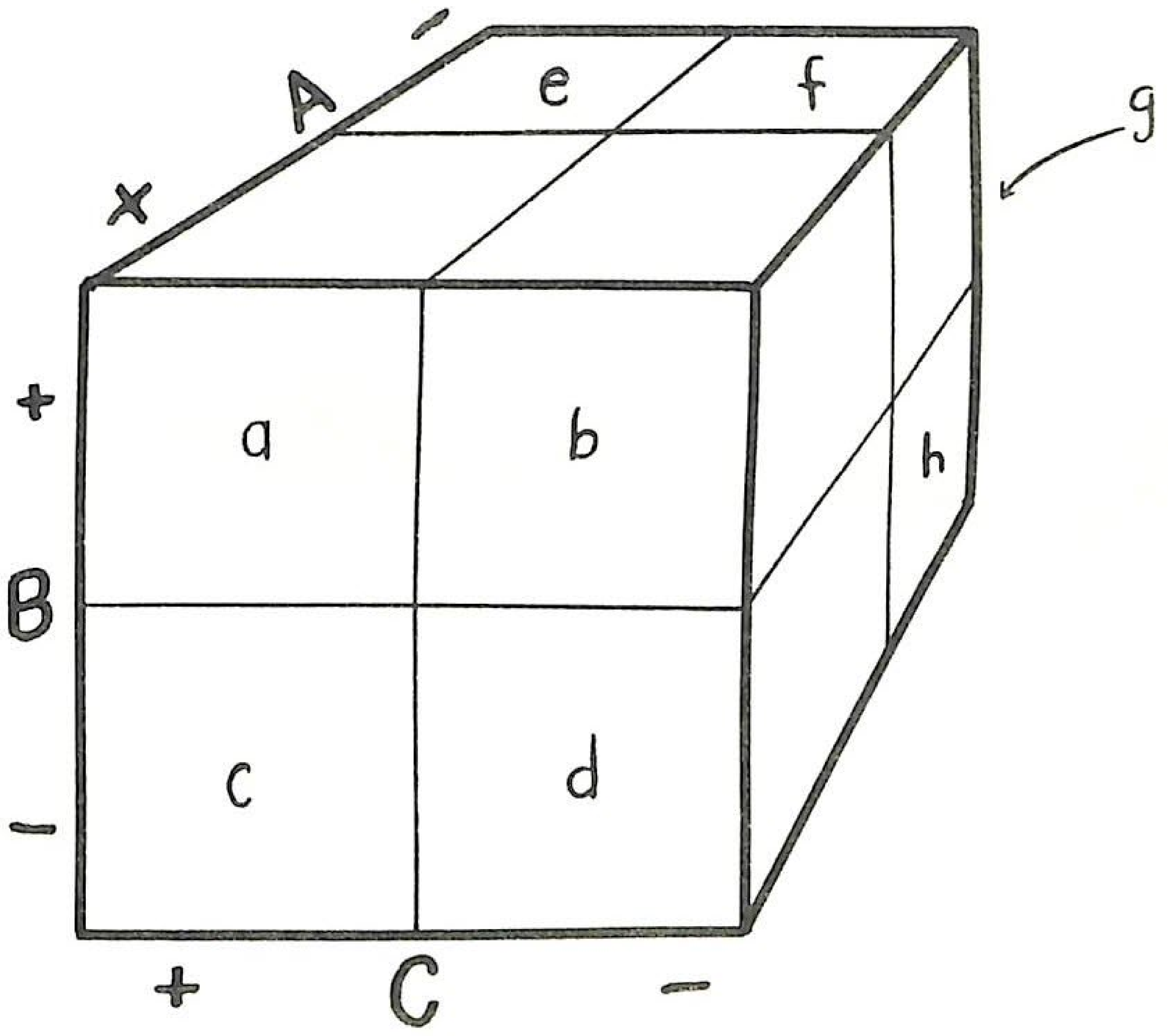
Item 62

| | Positive Responses | Negative Responses | Total |
|---|---|---|---|
| Rank 9 | 14 | 18 | 32 |
| Rank 8 | 31 | 14 | 45 |
| Total | 45 | 32 | 77 |

$\chi^2 = 3.885$

FIGURE 5

2 × 2 × 2 CONTINGENCY TABLE HOUSING THE FREQUENCIES OF THE JOINT OCCURRENCES AND NON-OCCURRENCES OF THREE ATTRIBUTES



fourth cell is determined by the marginal total of the positive responses to A.

The right face of the cube holds the four cells, b, d, f, and h, which house the frequencies of the negative responses to C. Two of these cells, b and d, have already been filled; hence the arbitrary choice of the frequency of either, but not both, f or h, is possible, the other being determined by the total number of negative responses to C. This uses one more degree of freedom.

The three cells, a, b, and f, of the upper face of the cube have now been determined, hence the frequency in cell e is determined by the total number of positive responses to B. Similarly, the frequency in g is determined by the total number of negative responses to B, cells c, d, and h having already been determined.

The solution above was simplified by first determining three cells in the same plane. However, it can be shown with simple algebra that there are only four degrees of freedom regardless of the order in which the individual cells are filled.

## APPENDIX

The items used in the inventory are listed below in order of the popularity of the "positive" response, and grouped to form the compound items of each scale. Each item is preceded by: 1) the number of the item in the inventory, 2) the proportion of individuals giving "positive" responses to the item, and 3) the range of responses designated as positive. The following notation is used to designate responses:

A - strongly agree
B - moderately agree
C - slightly agree
D - undecided or no response
E - slightly disagree
F - moderately disagree
G - strongly disagree

### Prudent-Theoretic Scale

A

| | | | |
|---|---|---|---|
| 68 | .31 | D-G | In science fiction stories, I like the ones with interesting scientific theories that hang together, even if they're not completely true, better than those about the social problems of space settlements. |
| 49 | .38 | D-G | In studying about the building of the pyramids, I should be more interested in the engineering feat than in the class structure and economy of Egypt which made such magnificent display possible. |
| 7 | .41 | A-D | Fabulous IBM machines are used to calculate insurance rates. However, data must be fed into the machines. The interviewing techniques for collecting data interest me more than an explanation of how the machines work. |
| 12 | .42 | A-E | I would rather teach science than do research. |

B

| | | | |
|---|---|---|---|
| 16 | .50 | D-G | In our complex industrial civilization a young person should specialize early and stick to it. |
| 54 | .61 | C-G | When I see an article about "electronic brains" I am more interested in finding out how they work than what their uses are. |
| 40 | .61 | A-E | I should rather read and be able to understand William Shakespeare's Hamlet than Michael Faraday's Experimental Researches in Electricity. |

| | | | |
|---|---|---|---|
| 44 | .62 | A-E | I like history and civics much better than science and mathematics. |

C

| | | | |
|---|---|---|---|
| 1 | .72 | A-E | I am more interested in finding out how TV has affected people's taste than in finding out how TV works. |
| 35 | .77 | C-G | I am more interested in following newspaper reports on the recent discoveries regarding "negative matter" than on the developments in racial integration in the schools. |
| 30 | .83 | B-G | I would rather study algebra than history even though algebra seems to be almost totally unrelated to any other subject. |
| 59 | .85 | B-G | If I had an hour to wait for a train I should more likely read The Scientific American than The Atlantic Monthly. |

## Prudent-Immediate Scale

A

| | | | |
|---|---|---|---|
| 23 | .30 | F-G | In a social studies course I would rather have the reasons why the U. S. didn't join the League of Nations explained to me than try to figure it out. |
| 36 | .43 | A-C | I should prefer the live theater to movies if they were the same price. |
| 65 | .47 | E-G | Sometimes when a fellow is out with the gang, he pretty well has to do a few things he knows he really shouldn't. |
| 71 | .47 | E-G | Mercy killing should be legalized for cases of extreme suffering where there is no hope for cure. |

B

| | | | |
|---|---|---|---|
| 19 | .53 | D-G | When talking with my friends in the evening I'd rather talk about people we know and have fun with than talk about religion or philosophy. |
| 4 | .53 | D-G | The opinion of friends helps more than reading in making up my mind. |
| 26 | .70 | A-E | The foreign policy of our government should be based on high moral principles even though this may entail a loss of strategic power or prestige. |
| 39 | .74 | A-D | I frequently think about the reasons for other people's misbehavior instead of reacting with irritation. |

C

| | | | |
|---|---|---|---|
| 9 | .76 | D-G | I never worry about how things are going to work out—they usually seem to take care of themselves. |
| 60 | .84 | D-G | An impulsive person is warm and sincere; one who analyzes his emotions is cold and "phony." |
| 51 | .84 | D-G | A business man should make his decisions strictly according to the interests of his business. He should not worry about what happens nationally to wages and prices. |
| 45 | .86 | A-D | Science has definitely not been able to show that colored races are inferior to white races. |

## Prudent-Aesthetic Scale

### A

| | | | |
|---|---|---|---|
| 64 | .29 | F-G | Nobody should be allowed to cut down the redwood forests and turn them into lumber. |
| 34 | .31 | E-G | If pay, housing, etc., were equal, I should like the work of a forest ranger better than that of a minister. |
| 57 | .45 | F-G | I should rather take a "shop" course than a "world history" course. |
| 48 | .46 | A-B | If I were a musician the thing I should like best about it would be getting across to the audience the basic idea of the composer. |

### B

| | | | |
|---|---|---|---|
| 3 | .49 | A-B | I find paintings interesting when I am able to see how they represent the artists' attitudes toward life. |
| 29 | .50 | F-G | Art should be appreciated intuitively. Analysis destroys its beauty. |
| 70 | .54 | F-G | I like to watch a big house afire. |
| 20 | .59 | F-G | To my way of thinking, the need to keep a city beautiful to look at is the most important argument in favor of smog control. |

### C

| | | | |
|---|---|---|---|
| 52 | .68 | F-G | Because they need to get close to life, artists are entitled to special consideration if they treat lightly the ties of marriage. |
| 8 | .69 | F-G | Medical experiments using live animals are cruel and inhuman. |
| 27 | .81 | E-G | Visiting a foreign country I would want to see the pageantry and architecture so I would not be interested in knowing in advance about their customs and history. |
| 42 | .82 | A-B | Instead of developing expensive tastes, what I would like most to get from my education is either a purpose for my life or an affirmation of my present purposes. |

## Theoretic-Immediate Scale

### A

| | | | |
|---|---|---|---|
| 31 | .34 | F-G | When I'm studying math or science it is refreshing to take frequent breaks watching TV or talking with a friend. |
| 10 | .35 | A-E | When I'm watching a movie I sometimes lose track of the plot because I'm wondering how the lighting and stage effects are worked. |
| 41 | .44 | A-D | I can visualize myself reading a paper at a scientific society meeting but not chatting socially in the corridor while a meeting is in progress. |
| 22 | .45 | A-C | I usually like to do math problems alone rather than discuss them with others. |

### B

| | | | |
|---|---|---|---|
| 6 | .47 | D-G | School mathematics courses should concentrate more on practical consumer and business training. |
| 53 | .55 | D-G | I would sooner have a big living room for parties than have a workroom for hobbies. |

| | | | |
|---|---|---|---|
| 2 | .58 | A-D | If I were employed by a company manufacturing chemicals, I would rather stay in research than become a company executive, so long as the loss in pay was not too great. |
| 62 | .63 | A-E | I prefer a science class that is run along fairly formal lines so that I can avoid the distractions arising from personal entanglements with other members of the class. |

C

| | | | |
|---|---|---|---|
| 69 | .71 | A-D | I enjoy working hard at a science project even if others don't recognize my accomplishment. |
| 14 | .74 | C-G | I should rather be elected to the Student Council than be selected as an honor student in science. |
| 46 | .75 | A-D | I don't like being interrupted while I'm doing laboratory experiments by friends who feel like talking. |
| 58 | .76 | C-G | I would rather be known as the writer of a social column published in many papers than as the Director of an astronomical observatory. |

## Theoretic-Aesthetic Scale

A

| | | | |
|---|---|---|---|
| 66 | .31 | F-G | If I were interested in studying flowers, I should be attracted chiefly by the beauty of the flowers. Comparing the structure of different kinds of flowers would not interest me much. |
| 17 | .42 | A-B | I prefer chess to checkers. |
| 56 | .46 | C-G | When you go on an automobile trip it is much more fun to pick places to stay as you go along rather than writing ahead for reservations. |
| 33 | .49 | E-G | I never wonder how the time is going when I'm painting a room or sawing firewood like I do when I'm studying math or physics. |

B

| | | | |
|---|---|---|---|
| 61 | .56 | E-G | Chemistry experiments are fun to watch so long as there are plenty of explosions and color changes. |
| 47 | .61 | E-G | A person should throw himself into life with vitality—the scientists' reflection on how things work is a wet blanket on the spontaneous pleasures of affection. |
| 38 | .65 | A-C | When I was little I liked erector sets more than tops. |
| 25 | .69 | A-C | It's a sloppy sailor who'll let his sails flap while he basks in the sun and breathes the crisp salty wetness of the air; the keen sailor watches the wind, studies the tides, and understands details of the rigging. |

C

| | | | |
|---|---|---|---|
| 28 | .71 | D-G | Scientists destroy much of the beauty of nature when they explain away its mysteries. |
| 21 | .78 | E-G | The forces of nature are subjects for wonder and awe—not analysis. |
| 13 | .79 | E-G | "Hotrod" racing would be fun if you didn't have to know about and work on motors. |
| 72 | .81 | A-C | When I look at the stars at night I sometimes meditate on the way the universe works. |

## Aesthetic-Immediate Scale

### A

| | | | |
|---|---|---|---|
| 24 | .37 | A-E | I would rather go sailing by myself than watch a football game. |
| 67 | .39 | A-D | If I had the necessary athletic prowess I should prefer to excel in the cross-country marathon than in football. |
| 11 | .41 | A-D | I like to ride alone. The feel of a good horse under me, his strength and his rhythm, more than make up for the lack of fellow riders. |
| 43 | .44 | C-G | Abilities of sign writers are different from the abilities of men who run a sign business. If I had the ability to do either, I would rather learn to run the business than paint the signs. |

### B

| | | | |
|---|---|---|---|
| 37 | .52 | A-E | I spend more of my free time on hobbies like stamp collecting, woodwork, etc., than going to parties or entertaining friends. |
| 18 | .53 | C-G | Sailing on a boat would be fun with a group of people, but I don't think I'd care much for it by myself. |
| 50 | .61 | A-E | I like Dixieland jazz better than "rock and roll." |
| 63 | .64 | C-G | A girl should wear sweaters and "pearls" or whatever most of the girls are wearing rather than conspicuous hand-made jewelry. |

### C

| | | | |
|---|---|---|---|
| 32 | .71 | A-F | I should prefer to be a machinist rather than a salesman. |
| 5 | .74 | A-F | I enjoy swimming in the ocean by myself, or, for safety, with a companion or life guard, more than swimming in a pool. |
| 15 | .79 | C-G | I think I should enjoy Longfellow's poem "Evangeline" more if it were told as a love story in modern prose. |
| 55 | .86 | B-G | When I am buying clothes I pay less attention to the ones that don't show than to the ones that do show. |

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 1, September 1959)

# STUDIES IN SOCIAL PERCEPTION: WORD PRODUCTIVITY

FRANK J. ESTVAN*
University of Wisconsin

THIS IS A study of the length of children's oral responses to a series of pictures. It is one phase of research dealing with children's social perception in which picture projection was employed. Word Productivity is of concern to workers in the field of projective testing who are faced with the problem of interpreting various measures of productivity. It has also been studied by specialists in human development and language arts whose findings seem to indicate that urban children are more verbal than rural, that girls excel boys in language development, and that children with high I.Q.'s are more verbally adept than those with less intelligence.

The present study is based on the assumption that behavior is caused, and that word count should, therefore, reveal something about the individual. Acceptance of this assumption would make tenable several explanations for the number of words produced in response to the series of pictures used in this research. One would be that Word Productivity is an index of interest or effort, sometimes referred to as subject cooperation. Another is that Word Productivity is a reflection of the meaningfulness of a situation to the individual. Third, Word Productivity may be an indicator of the pictorial quality or provocativeness of a picture irrespective of its content.

To test these hypotheses, comparisons have been made of the Word Productivity of rural and urban children, boys and girls, first and sixth grade pupils, as well as high I.Q. and low I.Q. groups. If differences among these groups can be explained in terms of what is already known about children or by a new set of constructs, the usefulness of Word Productivity as an aid in the investigation of social perception will have been established.

It must be noted that the verbal behavior being examined is not what an individual knows about words, but the number of words he uses to communicate ideas. This distinction is the same as that between knowing how well a child can read, and what he does read. The meaningfulness of the responses elicited in this study will be found in another source.[1]

## Source of Data[2]

The data analyzed in this report were obtained in Part I of the Social Perception Interview. A Life-Situation Picture Series consisting of 14 black and white 2 × 2 inch slides was projected on a 30 × 40 inch screen, the directive for each stimulus being "What story does this picture tell?" The subject's remarks were tape recorded and typewritten verbatim to yield a highly accurate record of his free-association responses.

Of the total of 88 subjects selected at random, half were drawn in equal proportions from the first and sixth grade children attending one-room schools in a rural Wisconsin county. The other 44 subjects were similarly selected from the first and sixth grade pupils attending the public elementary schools of a Wisconsin community of approximately 55,000 population located in a highly urbanized region. Sex was equated within each grade level for both rural and urban samples, thus providing eight subgroups of 11 subjects each. Intelligence groups were formed by selecting the two persons having the highest Intelligence Quotient in each of the eight subgroups to constitute the high-ability sample, and using similar procedures at the opposite extreme to form the low-ability group.

The determination of Word Productivity involved more than the mechanical counting of a succession of words. In many instances, consideration had to be given to the context in which an expression occurred. To standardize procedures, the following guides were established:

A. Counted as one word
   1. Any word contained in a standard English dictionary.
   2. Slang expressions or colloquialisms—cop, yeah, sorta', lotsa'.
   3. Contractions—isn't, there's, ain't.
   4. Compound words—right-handed.
   5. Abbreviations—YMCA, O.K., ABC's.
   6. Numbers below 100.

---

*All footnotes will be found at end of article.

7. Partial words when the context indicated their use as a complete word—"semi" (semi-trailer truck), "spose" (suppose), "veyor" (surveyor).
8. Repetitions when they were part of the growing perception of the subject—"boy...I see a boy" (each word counted), "and...and a girl."

B. Not counted
1. Sounds—ah, um, mm.
2. Incompleted words which occurred when the subject changed his mind after he had begun a word—"sa..." (sad), "peo..." (people).
3. Repetitions or stammering unrelated to thought processes—"and...and... and..." (latter two not counted), "a...a...a...a..." (latter three not counted).

In order to eliminate observer influence, no attempt was made to estimate the number of words missed in those few instances when responses were not audible on the tape recorder. Neither was the observer required to determine when the child had stopped talking about one life-situation picture and was signalling for the next, such remarks as "that's all," "O.K.," "I don't know any more" being included in the word count for the picture being viewed on the screen.

## Does Word Productivity Indicate Pupil Involvement?

This is the first and most basic question which needs to be answered. If the differences in Word Productivity among the pictures in the series are so small as to reflect the operation of chance, a lack of personal involvement on the part of the subjects could be inferred. This would tend to invalidate the data secured in the picture-story portion of the Social Perception Interview, and cast doubt on all the procedures which followed.

It is conceivable that the life-situation pictures might be equally stimulating in content or form of presentation to the subjects in this study. The fact that these situations vary in: a) social background representation (rural-urban, high status-low status, child-adult), b) social functions depicted (10 basic life activities), c) scope or complexity of situations (general environment-specific social function), and d) child-figure presentation (clearly defined or ambiguous) would seem to lessen such a possibility. For all these variables to average out among the 14 pictures in approximately equal word responses is not likely except that subjects react in a random manner rather than to the elements in each situation.

Table I presents the variance in Word Productivity of the 88 subjects to each of the 14 life-situation pictures. The test for homogeneity of variance establishes these differences at a very high level of confidence. This means that the variability in Word Productivity among the 14 pictures is too great to consider these responses as belonging to the same population. In short, the differences in Word Productivity among the pictures in the series would appear to be real and not simply due to chance. The subjects in this study were not responding to the Life-Situation Picture Series in random fashion.

## Is Word Productivity Related to Certain Pupil Abilities?

In considering the significance of Word Productivity, the possibility exists that the total number of words produced is simply an index of general intelligence or a more specific verbal factor. It may be that individuals have developed a generalized verbal behavior pattern, as far as quantity is concerned, which is more influential in determining Word Productivity than are the pictures themselves. The broader base of general intelligence must also be examined in this connection, because of the high relationship which usually prevails between scores on intelligence tests and tests of verbal abilities.

A high correlation between Word Productivity and intelligence or verbal facility would discount the meaningfulness of the subject's reactions to the life situations. The quantitative aspect of his responses could then be attributed to an underlying "native" ability in general or to a more specific verbal factor. If no relationship were found between these measures, one would have greater confidence in regarding differences in Word Productivity as being associated with differences in the meaningfulness that these situations have for elementary school children.

Measures of intelligence and verbal ability were obtained from the Stanford-Binet Intelligence Test, Form L. The Vocabulary score is the number of words for which correct definitions were given until six consecutive words were missed. These data were available for all 88 subjects included in the study, and may be regarded as a qualitative aspect of verbal behavior. The Word Fluency score is the number of words named in one minute and, thus, is a quantitative measure of verbal facility. The latter is Item 5 for Year X, and was attempted by 55 percent of the first-grade children and 73 percent of the sixth-grade children, thus providing sufficient data for analysis.

As is indicated in Table II, none of the correlations between Word Productivity and a) Mental Age, b) Intelligence Quotient, c) Vocabulary Score and d) Word Fluency Score are statistically significant for the first grade, the sixth grade, or for both groups combined. In every instance, however, the correlation of Vocabulary Score with Mental Age and Intelligence Quotient is signifi-

## TABLE I

## HOMOGENEITY OF VARIANCE IN WORD PRODUCTIVITY OF TOTAL GROUP FOR LIFE-SITUATION PICTURE SERIES

| Picture | | | | | |
|---|---|---|---|---|---|
| No. | Name | N | $\Sigma X$ | $\Sigma X^2$ | $s^2$ |
| 1 | Bedroom | 88 | 2619 | 125839 | 550.50 |
| 2 | Hovel | 88 | 3747 | 227797 | 784.50 |
| 3 | Dam | 88 | 3877 | 253489 | 950.35 |
| 4 | Factory | 88 | 3391 | 206103 | 867.05 |
| 5 | Mansion | 88 | 3989 | 253121 | 831.05 |
| 6 | Church | 88 | 2906 | 144502 | 557.90 |
| 7 | Farm | 88 | 3448 | 203506 | 786.29 |
| 8 | Swimming Hole | 88 | 3399 | 211741 | 924.77 |
| 9 | Capitol | 88 | 3296 | 192746 | 796.50 |
| 10 | City | 88 | 3438 | 205728 | 820.82 |
| 11 | Resort | 88 | 3474 | 211676 | 856.68 |
| 12 | Schoolroom | 88 | 3380 | 217738 | 1010.52 |
| 13 | Village | 88 | 3494 | 237508 | 1135.41 |
| 14 | Dock | 88 | 3929 | 287607 | 1289.50 |

$$F_{max} = \frac{s^2 max}{s^2 min} = \frac{1289.50}{550.50} = 2.34^{**} \qquad s^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}$$

For $k = 14$ and $n = 87$: $P(F_{max} > 1.80) = .01^{**}$

## TABLE II

## CORRELATION OF WORD PRODUCTIVITY AND MEASURES OF INTELLIGENCE AND VERBAL ABILITY

| | First Grade | | | Sixth Grade | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | |
| 1 | | | | | | | 1 |
| 2 | .106 | | | .034 | | | 2 |
| 3 | .140 | .985** | | .032 | .993** | | 3 |
| 4 | .178 | .741** | .739** | -.119 | .683** | .700** | 4 |
| 5 | .038 | .254 | .325* | .210 | .100 | .090 | 5 |

| Total Group | | | |
|---|---|---|---|
| 1 | 2 | 3 | |
| | | | 1 |
| .144 | | | 2 |
| .097 | .482** | | 3 |
| .130 | .931** | .411** | 4 |
| .120 | .569** | -.217* | 5 |

Key:

1— Word Productivity
2— Mental Age
3— Intelligence Quotient
4— Vocabulary Score
5— Word Fluency Score[1]

[1] N = 24 first grade and 32 sixth grade instead of 44 for each grade as for other measures.

* P < .05
**P < .01

cant beyond the one percent level of confidence. For the total group of 56 subjects who responded to the item, the correlation of Word Fluency Score with Mental Age and Intelligence Quotient is likewise significantly different from zero. This is also true with respect to Word Fluency Score and Intelligence Quotient in the First Grade group.

The fact that Word Productivity is not related to level or rate of intelligence as well as to qualitative or quantitative measures of verbal ability, would support the conclusion that word responses to the Life-Situation Picture Series reflect something other than these measures. The Social-Perception Interview, from the standpoint of word productivity, can make no claim to being another form of intelligence test or test of vocabulary.

## Is Word Productivity Related to Characteristics of the Instrument?

Differences in Word Productivity might also be explained in terms of certain characteristics of the instrument rather than being attributed to differences in the meaningfulness of the life situations to subjects. It could be that while the subject-matter of a picture is interesting to elementary-school children, its pictorial treatment has little appeal. There is also the possibility that the placement of a picture in the series will effect the length of verbal responses elicited. Word Productivity of the beginning and latter pictures in the series may simply be indications of the need for a "warming-up" period or fatigue due to the length of the series.

Pictorial Representation—One way to examine the first question is to determine the relationship in Word Productivity among the various pictures in the series. A significant negative correlation between one picture and the others in the series would call for extended analysis to discover the factors which might explain such deviate responses. These could be matters of form or content. The intercorrelations are positive and statistically significant, thus lending a high degree of confidence in a relationship that is not zero. Of the 91 correlations, 30 are +.80 or higher, the highest being between Picture No. 10 (City) and Picture No. 13 (Village) which are the most clear-cut paired-comparisons in the series. Except for Picture No. 1 (Bedroom), each picture correlates +.80 or higher with one or more pictures; Picture No. 12 (Schoolroom) correlating with nine other pictures at this level of relationship.

The generally high positive correlation in Word Productivity among the 14 life-situation pictures indicates that no one picture stands apart from the rest of the series because of extremely high or low stimulating properties. Although the pictures in this series vary in their power to elicit word responses from elementary school children, this difference is not so extreme in any one picture as to question its suitability in content or pictorial qualities.

Picture Sequence—The fact that the first picture in the series is the only one which did not correlate as high as +.80 with others in the series is one reason for investigating the relationship between Word Productivity and Picture Sequence. A significant negative correlation could mean that subjects gained in confidence and interest and, hence, increased in Word Productivity as they proceeded through the series—or that the pictures toward the end of the series were more provocative. A high positive correlation, on the other hand, could mean that subjects became fatigued or lost interest as they proceeded, that the latter pictures in the series were less stimulating, or that subjects became more direct and efficient in expressing themselves as they gained practice. A lack of relationship between Word Productivity and Picture Sequence would indicate the absence of systematic influences such as "warming up," fatigue, or practice effects, and an evenness in the stimulating quality of the life-situation pictures.

Table IV indicates that for the total group of 88 subjects there is no relationship between Word Productivity and the order in which pictures are presented. This is also true for the comparative groups listed in Table V with the exception of the High Intelligence group. In this case, both rho and Kendall's tau are significantly different from zero at the 4 percent level of confidence. If the pictures at the end of the series were more stimulating, we would expect an increase in Word Productivity from all groups. In the absence of this tendency, a more plausible explanation is that bright children profit greatly from experience and, therefore, reach higher levels of performance as they progress through the series. For elementary-school children, in general, it is clear that differences in the Word Productivity of various pictures are not related to their position in the series.

It is interesting to note in Table V that the rankings for Picture No. 1 (Bedroom) and Picture No. 6 (Church) differ by only one point among the eight subgroups, and that they rank 14th and 13th respectively. Whether scenes of a child in bed and people going to church seem so obvious to children as to need no prolonged comment or whether they are less interesting or meaningful than the other pictures can only be surmised. For each of the remaining 12 pictures, however, the rankings made by the comparative groups differ from 4 to 9 steps, the average rank-difference being 7.

## Is Word Productivity Related to Differences in Social Background and Intelligence?

Meaningfulness is a function of the kind of experiences which an individual has, and his ability to learn from these experiences. Accordingly, it

## TABLE III

## INTERCORRELATIONS IN WORD PRODUCTIVITY OF LIFE-SITUATION PICTURES

| | Life-Situation Picture | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1 | | | | | | | | | | | | | | |
| 2 | .70 | | | | | | | | | | | | | |
| 3 | .62 | .84 | | | | | | | | | | | | |
| 4 | .50 | .77 | .78 | | | | | | | | | | | |
| 5 | .53 | .81 | .83 | .84 | | | | | | | | | | |
| 6 | .59 | .78 | .83 | .77 | .81 | | | | | | | | | |
| 7 | .60 | .80 | .82 | .77 | .78 | .80 | | | | | | | | |
| 8 | .58 | .71 | .72 | .79 | .76 | .78 | .80 | | | | | | | |
| 9 | .58 | .78 | .76 | .72 | .73 | .66 | .76 | .71 | | | | | | |
| 10 | .58 | .80 | .78 | .71 | .79 | .81 | .79 | .78 | .84 | | | | | |
| 11 | .49 | .73 | .73 | .73 | .82 | .71 | .84 | .79 | .85 | .83 | | | | |
| 12 | .61 | .76 | .80 | .75 | .79 | .81 | .82 | .83 | .82 | .84 | .84 | | | |
| 13 | .53 | .79 | .78 | .71 | .80 | .86 | .79 | .77 | .76 | .89 | .79 | .84 | | |
| 14 | .47 | .69 | .69 | .73 | .76 | .73 | .72 | .76 | .76 | .82 | .78 | .82 | .77 | |

For 86 df: $P(r > .210) = .05$
$P(r > .274) = .01$

## TABLE IV

## CORRELATION OF WORD PRODUCTIVITY AND PICTURE SEQUENCE FOR TOTAL GROUP

| Picture No. | Words | | Rank |
|---|---|---|---|
| | Total | Mean | |
| 1......... | 2619 | 29.76 | 14 |
| 2......... | 3747 | 42.58 | 4 |
| 3......... | 3877 | 44.06 | 3 |
| 4......... | 3391 | 38.53 | 10 |
| 5......... | 3989 | 45.33 | 1 |
| 6......... | 2906 | 33.02 | 13 |
| 7......... | 3448 | 39.18 | 7 |
| 8......... | 3399 | 38.62 | 9 |
| 9......... | 3296 | 37.45 | 12 |
| 10......... | 3438 | 39.07 | 8 |
| 11......... | 3474 | 39.48 | 6 |
| 12......... | 3380 | 38.41 | 11 |
| 13......... | 3494 | 39.70 | 5 |
| 14......... | 3929 | 44.65 | 2 |
| Total | 48387 | 39.28 | |

rho = -.147

## TABLE V

## CORRELATION OF WORD PRODUCTIVITY AND PICTURE SEQUENCE FOR COMPARATIVE GROUPS

| Picture | Community | | Sex | | Grade | | Intelligence | |
|---|---|---|---|---|---|---|---|---|
| | Rural | Urban | Boy | Girl | 1st | 6th | High* | Low* |
| 1......... | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 13 |
| 2......... | 2 | 4 | 4 | 4 | 1 | 8 | 8 | 2 |
| 3......... | 4 | 2 | 2 | 3 | 2 | 3 | 4 | 9 |
| 4......... | 10 | 10 | 5 | 12 | 8 | 9 | 12 | 3.5 |
| 5......... | 1 | 3 | 1 | 2 | 3 | 2 | 7 | 1 |
| 6......... | 13 | 13 | 13 | 13 | 12 | 13 | 13 | 12 |
| 7......... | 6 | 12 | 7 | 10 | 4 | 11 | 10 | 11 |
| 8......... | 9 | 11 | 6 | 11 | 13 | 4 | 11 | 5 |
| 9......... | 12 | 8 | 12 | 8 | 10 | 12 | 5 | 14 |
| 10......... | 5 | 7 | 11 | 6 | 5 | 10 | 3 | 10 |
| 11......... | 7 | 9 | 8 | 7 | 9 | 5 | 2 | 7 |
| 12......... | 11 | 6 | 10 | 9 | 11 | 6 | 6 | 8 |
| 13......... | 8 | 5 | 9 | 5 | 6 | 7 | 9 | 6 |
| 14......... | 3 | 1 | 3 | 1 | 7 | 1 | 1 | 3.5 |
| rho | -.033 | -.275 | .112 | -.293 | .143 | -.332 | -.556** | -.041 |

* N = 16 instead of 44 as for other groups.

**For rho = — .556
t = — 2.321
P < .04

**Kendall's tau = — .429
z = — 2.080
P = .04

**For computing rho, t, and Kendall's tau see: Maurice G. Kendall, Rank Correlation Methods (London: Charles Griffin and Co., Ltd., 1948), Chaps. I and IV.

would be reasonable to expect that children having different social backgrounds such as those associated with rural-urban living, sex, and age (grade) would respond differently to a variety of scenes of American life. For the same reason, the reactions of bright children might differ from those who are less intelligent.

*Community and Sex Differences*—To determine whether Word Productivity is related to type of community lived in and to sex, analyses were made of the four first-grade subgroups (1RB, 1RG, 1UB, 1UG) and another of the four sixth-grade subgroups (6RB, 6RG, 6UB, 6UG).[3] Tests of homogeneity of variance for each picture, summarized in Table VI, indicate that with one exception in the first grade, and two in the sixth, the subgroups did not differ significantly in the variability of their responses. The first-grade urban girls contributed disproportionately to the variance for Picture No. 14 (Dock) and the sixth-grade rural boys accounted for the great variability in responses to Picture No. 4 (Factory) and to Picture No. 8 (Swimming Hole). In all three instances, sex produced greater variance than community lived in. The results of the analysis of variance for the first-grade subgroups and for the sixth-grade subgroups are presented in Table VII, and the main effects variance in Table VIII. In none of the 56 analyses (14 pictures × 2 variances × 2 grade groups) was the F ratio statistically significant. This is convincing evidence that the locality in which the individual lives and sex are not related to the mean number of words given in response to the life-situation pictures.

*Grade Differences*—A comparison of the Word Productivity of the 44 first-grade subjects and the 44 sixth-grade subjects is presented in Table IX. Except for Picture No. 1 (Bedroom) and Picture No. 2 (Hovel), the mean number of words produced by the sixth-grade pupils was greater than that for the first. In two cases, Picture No. 5 (Mansion) and Picture No. 14 (Dock), this difference was great enough to be considered significant rather than due to sampling fluctuations. In both groups, the Mansion scene ranked high in Word Productivity, the sixth grade placing it second and the first grade ranking it third. Whereas the Dock scene ranked highest in the sixth-grade group, it placed only seventh in the first grade.

It would appear that both groups have come to realize the significance of living in wealthy circumstances, and that the difference between the first and sixth grade on the Mansion picture is one of degree due to the fact that sixth grade children have had a longer period of acculturation. To understand why the greatest difference between the grade groups should appear with reference to the Dock scene, it is important to note that this picture was planned to depict human relations on a world-wide scale. That sixth grade children were more stimulated to talk about world relationships than were first graders, fits in with the commonly held notion that children proceed from the "near" to the "far" in their development of interests.

*Intelligence Differences*—In Table X, the Word Productivity of 16 high I. Q. subjects and 16 low I. Q. pupils are compared for each of the life-situation pictures. In every case, the mean Word Productivity of the former group exceeds that of the latter. For six of the 14 pictures, the difference is statistically significant and not due to chance. This is considerable evidence to the effect that bright children, irrespective of where they live, their sex, or grade are more stimulated to talk about certain life-situation pictures than are a comparable group of children who differ in being below average in intelligence.

Why bright children should be more productive on the City (No. 10) and Farm (No. 7) pictures in the community pattern and to the Hovel (No. 2) and Resort (No. 11) in the social status block of pictures is not immediately apparent. The picture of a Dam (No. 3) and that of the Capitol (No. 9), however, are two of the three *adult* pictures designed to contrast age differentials in social experience. It is, therefore, not difficult to understand why children of high intelligence would be more stimulated by scenes depicting the conservation of natural resources as in the Dam and of citizenship values as reflected in the Capitol building than would be the case with less intelligent children.

### Are Extreme Deviations in Word Productivity for Each Individual Related to Differences in Social Background and Intelligence?

In the foregoing analyses, Word Productivity has been examined from the point of view that significance is in direct proportion to magnitude. Significant behavior, however, may be expressed by *avoidance* as well as *attraction*, and either may be prompted by the same underlying causes. With reference to Word Productivity, this would be indicated by an extremely meager word count or by an exceptionally productive response. A determination of deviant behavior in these terms would have to take into account the wide range of individual differences in general level of Word Productivity. A word count of 63, for example, might indicate a high response for one individual, and average or below average productivity for another person. Hence, the requirement that each individual's word responses constitute the basis for this type of analysis.

To determine what is a deviant response for an individual, the assumption must be made that word responses to the 14 pictures are normally distributed. It is then possible to determine mean Word Productivity and variability in word responses for the series. If the responses beyond ±1 SD from the individual's mean Word Productivity are des-

TABLE VI

HOMOGENEITY OF VARIANCE IN WORD PRODUCTIVITY OF FIRST AND SIXTH GRADE SUBGROUPS FOR EACH PICTURE

| Picture | 1st Grade $F_{max}$ | 6th Grade $F_{max}$ |
|---|---|---|
| 1......... | 2.69 | 2.70 |
| 2......... | 2.17 | 2.23 |
| 3......... | 1.66 | 1.59 |
| 4......... | 4.45 | 10.17** |
| 5......... | 2.97 | 2.93 |
| 6......... | 1.37 | 3.06 |
| 7......... | 5.04 | 1.91 |
| 8......... | 2.74 | 9.05** |
| 9......... | 2.20 | 2.65 |
| 10......... | 3.19 | 2.72 |
| 11......... | 2.63 | 2.84 |
| 12......... | 4.95 | 2.64 |
| 13......... | 3.69 | 2.60 |
| 14......... | 7.32* | 2.18 |

For k = 4 and N = 10: *$P(F_{max} > 5.67) = .05$
**$P(F_{max} > 8.60) = .01$

## TABLE VII

### ANALYSIS OF VARIANCE IN WORD PRODUCTIVITY OF FIRST AND SIXTH GRADE SUBGROUPS FOR EACH PICTURE

| | First Grade | | | Sixth Grade | | | |
|---|---|---|---|---|---|---|---|
| Picture | Between | Within | F | Between | Within | F | Picture |
| 1......... | 100.88 | 828.46 | .12 | 123.72 | 351.77 | .35 | .........1 |
| 2......... | 58.36 | 1242.05 | .05 | 31.66 | 457.55 | .07 | .........2 |
| 3......... | 195.90 | 1249.52 | .16 | 540.43 | 746.63 | .72 | .........3 |
| 4......... | 313.54 | 972.25 | .32 | 569.51 | 829.85 | .69 | .........4 |
| 5......... | 51.60 | 744.11 | .07 | 148.27 | 954.55 | .16 | .........5 |
| 6......... | 25.64 | 714.03 | .04 | 44.76 | 494.23 | .09 | .........6 |
| 7......... | 406.45 | 948.39 | .43 | 146.87 | 713.93 | .21 | .........7 |
| 8......... | 6.70 | 713.90 | .01 | 593.60 | 1167.55 | .51 | .........8 |
| 9......... | 190.58 | 987.44 | .19 | 703.79 | 653.70 | 1.08 | .........9 |
| 10......... | 271.21 | 1038.50 | .26 | 221.24 | 700.53 | .32 | .........10 |
| 11......... | 59.11 | 829.81 | .07 | 49.48 | 974.30 | .05 | .........11 |
| 12......... | 358.81 | 1002.08 | .36 | 619.54 | 1061.88 | .58 | .........12 |
| 13......... | 573.97 | 1293.52 | .44 | 232.97 | 1094.18 | .21 | .........13 |
| 14......... | 481.72 | 1120.15 | .43 | 205.52 | 1469.38 | .14 | .........14 |

For 3 and 40 df: *$P(F>2.84) = .05$
**$P(F>4.31) = .01$

TABLE VIII

ANALYSIS OF MAIN EFFECTS VARIANCE IN WORD PRODUCTIVITY OF FIRST AND SIXTH GRADE SUBGROUPS FOR EACH PICTURE

| | First Grade | | | Sixth Grade | | | |
|---|---|---|---|---|---|---|---|
| Picture | Between | Within | F | Between | Within | F | Picture |
| 1......... | 302.64 | 828.46 | .37 | 371.16 | 351.77 | 1.06 | .........1 |
| 2......... | 175.09 | 1242.05 | .14 | 94.98 | 457.55 | .21 | .........2 |
| 3......... | 587.70 | 1249.52 | .47 | 1621.28 | 746.63 | 2.17 | .........3 |
| 4......... | 940.61 | 972.25 | .97 | 1708.54 | 829.85 | 2.06 | .........4 |
| 5......... | 154.80 | 744.11 | .21 | 444.82 | 954.55 | .47 | .........5 |
| 6......... | 76.91 | 714.03 | .11 | 134.27 | 494.23 | .27 | .........6 |
| 7......... | 1219.34 | 948.39 | 1.29 | 440.61 | 713.93 | .62 | .........7 |
| 8......... | 20.09 | 713.90 | .03 | 1780.80 | 1167.55 | 1.53 | .........8 |
| 9......... | 571.73 | 987.44 | .58 | 2111.37 | 653.70 | 3.23 | .........9 |
| 10......... | 813.61 | 1038.50 | .78 | 663.73 | 700.53 | .95 | .........10 |
| 11......... | 177.34 | 829.81 | .21 | 148.43 | 974.30 | .15 | .........11 |
| 12......... | 1076.44 | 1002.08 | 1.07 | 1858.62 | 1061.88 | 1.75 | .........12 |
| 13......... | 1721.91 | 1293.52 | 1.33 | 698.91 | 1094.18 | .64 | .........13 |
| 14......... | 1445.16 | 1120.15 | 1.29 | 616.55 | 1469.38 | .42 | .........14 |

For 1 and 40 df: $*P(F>4.08) = .05$
$**P(F>7.31) = .01$

## TABLE IX

## WORD PRODUCTIVITY OF FIRST AND SIXTH GRADE GROUPS FOR EACH PICTURE

| Picture | First Grade |  | Sixth Grade |  |  |  |
|---|---|---|---|---|---|---|
|  | Mean | SD | Mean | SD | $SE_{M_1-M_6}$ | t |
| 1......... | 30.14 | 27.57 | 29.39 | 18.12 | 5.03 | .149 |
| 2......... | 42.64 | 33.66 | 42.52 | 20.45 | 6.01 | .019 |
| 3......... | 41.39 | 33.90 | 46.73 | 26.75 | 6.59 | .811 |
| 4......... | 35.70 | 30.09 | 41.36 | 28.16 | 6.29 | .900 |
| 5......... | 38.80 | 26.08 | 51.86 | 29.63 | 6.02 | 2.171* |
| 6......... | 33.00 | 25.51 | 33.05 | 21.27 | 5.07 | .009 |
| 7......... | 37.48 | 29.83 | 40.89 | 25.67 | 6.00 | .568 |
| 8......... | 32.41 | 25.48 | 44.84 | 33.19 | 6.38 | 1.948 |
| 9......... | 34.14 | 30.18 | 40.77 | 25.34 | 6.01 | 1.104 |
| 10......... | 37.00 | 31.03 | 41.14 | 25.53 | 6.13 | .675 |
| 11......... | 34.66 | 27.54 | 44.30 | 29.82 | 6.19 | 1.557 |
| 12......... | 33.16 | 30.58 | 43.66 | 31.74 | 6.72 | 1.562 |
| 13......... | 36.59 | 34.86 | 42.82 | 31.79 | 7.19 | .866 |
| 14......... | 36.02 | 32.42 | 53.27 | 36.74 | 7.47 | 2.309* |

For 86 df: $*P(t > 1.99) = .05$
$**P(t > 2.64) = .01$

TABLE X

WORD PRODUCTIVITY OF HIGH AND LOW INTELLIGENCE GROUPS FOR EACH PICTURE

| Picture | High I. Q. | | Low I. Q. | | $SE_{M_H-M_L}$ | t |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | |
| 1......... | 35.63 | 19.09 | 26.63 | 26.75 | 8.22 | 1.095 |
| 2......... | 52.50 | 21.14 | 34.44 | 28.30 | 8.83 | 2.045* |
| 3......... | 55.81 | 29.97 | 29.88 | 27.10 | 10.10 | 2.567* |
| 4......... | 44.94 | 29.74 | 34.38 | 37.09 | 11.89 | .889 |
| 5......... | 52.75 | 24.93 | 40.88 | 36.57 | 11.07 | 1.073 |
| 6......... | 37.13 | 21.07 | 28.50 | 25.55 | 8.28 | 1.042 |
| 7......... | 51.44 | 31.84 | 29.13 | 28.37 | 10.66 | 2.093* |
| 8......... | 50.63 | 29.00 | 34.13 | 50.40 | 14.54 | 1.135 |
| 9......... | 53.75 | 39.61 | 25.63 | 24.64 | 11.66 | 2.412* |
| 10......... | 56.75 | 31.58 | 29.81 | 30.34 | 10.95 | 2.460* |
| 11......... | 59.81 | 41.68 | 30.31 | 30.75 | 12.95 | 2.278* |
| 12......... | 53.56 | 39.11 | 30.19 | 37.01 | 13.46 | 1.736 |
| 13......... | 51.88 | 30.39 | 30.63 | 33.09 | 11.23 | 1.892 |
| 14......... | 60.63 | 37.73 | 34.38 | 42.43 | 14.20 | 1.849 |

For 30 df: $*P(t > 2.042) = .05$
$**P(t > 2.750) = .01$

ignated as exceptional for him, approximately 32 percent or an average of 4.4 of his picture responses would be included in the negative and positive tails of the distribution. If cut-off points are established at ±2 SD from the mean, somewhat less than 5 percent or .64 pictures would be singled out. (In actual practice, the latter resulted in the identification of such unusual responses in one of two subject's responses.)

Total Group—The number of instances when the Word Productivity for a picture was beyond ±1 SD and ±2 SD from each individual's mean is shown for the total group of 88 subjects in Table XI. Although there is no appreciable difference between the total number of very limited responses and those showing great productivity at the ±1 SD level, the difference is highly significant for the ±2 SD scores. This tendency for extremely marked variability to be in the direction of over- rather than under-production is, therefore, one which may be held with a high degree of confidence. Its logic stems from the fact that whereas the zero limitation applies to the negative end of the distribution, there is no ceiling for responses falling in the positive side of the scale.

At the ±1 SD level of variability, five pictures show a statistically significant difference between the number of times they evoked little or much productivity. In Picture No. 2 (Hovel), Picture No. 5 (Mansion), and Picture No. 14 (Dock), the direction is positive whereas in Picture No. 1 (Bedroom) and Picture No. 6 (Church) the direction is negative. This is in keeping with the findings presented in Table IV showing these pictures to rank 4, 1, 2, and 14, 13 respectively in Word Productivity.

Differences between the ±2 SD scores for each picture are not statistically significant or are too small to be tested. It is interesting to note, however, that except for Picture No. 3 (Dam), the Bedroom and Church scenes are the only ones for which such an extremely limited response was forthcoming. The Bedroom situation, however, is also among the highest producers at the ±2 SD level which is not the case for the Church scene. It should also be noted that Picture No. 12 (Schoolroom) is the only one which did not elicit at least one voluminous response from the 88 children concerned.

The above findings support the generalization that elementary-school children are more stimulated to talk about conditions of poverty (No. 2) and wealth (No. 5) as well as far away places intimated in the Dock scene (No. 14) and to be less stimulated by such immediate experiences as sleeping (No. 1) and going to church (No. 6). It must be remembered, however, that these can be held only as tentative conclusions for it is quite possible that another set of illustrations of the same life activities might result in different responses.

Comparative Groups—Analyses of the preceding data in terms of community, sex, grade, and intelligence comparisons are given in Table XII (±1 SD) and Table XIII (±2 SD). In no instance is the distribution of positive and negative deviations for two groups significantly different. Each exhibits the same general trends as for the total group reported in Table XI, or varies so little as to be considered a random fluctuation from these trends.

When these same data are compared in terms of the social-background patterns on which the Life-Situation Picture Series was designed, certain statistically significant differences appear in all but the rural-urban comparison (Table XIV). Girls were more productive on high social-status pictures, but boys responded better to low social-status situations (Table XV). Whereas first-grade children were less productive on adult scenes. sixth-grade pupils were less stimulated by pictures centered on child-experiences (Table XVI). The high intelligence group was more responsive to the adult pictures, and the low intelligence group responded more freely to the child-experience pictures (Table XVII).

That rural children would differ from urban children in their reactions to scenes portraying rural-urban differentials seems to be a reasonable expectation. Why such differences in Word Productivity failed to materialize cannot be accounted for at this time. It may be that this differential does not exist in Southeastern Wisconsin. The rural-urban groups being equated in sex and grade, no differences were expected in the reactions of these groups to the social status and age patterns, and none appeared.

Why boys should talk at such great length to low social-status pictures can be explained, in part, by the weighting produced by the Swimming Hole scene (No. 8) which might be more attractive to boys than to girls. The consistently greater response of the girls to the high social-status pictures is in keeping with developmental data bearing on the advanced social maturity of girls over that of boys. With this maturity could go an increased sensitivity to the status symbols depicted in the Mansion (No. 5) and Resort (No. 11) scenes. There being equal representation of rural-urban and first- and sixth-grade children in the sex groups, there would be no reason for anticipating differences in their responsiveness to the community block of situations and the age pattern. No sex differences were discovered in these two patterns.

That differences in grade and, hence, age should be reflected in different reactions to the child-adult pattern of life-situations can also be explained logically. As was indicated in the preceding discussion on grade differences in total Word Productivity, this is in keeping with present knowledge of the increasing span or breadth of

TABLE XI

PICTURES DEVIATING $\pm 1$ SD AND $\pm 2$ SD FROM INDIVIDUAL MEAN WORD PRODUCTIVITY FOR TOTAL GROUP

| Picture | SD | | $X^2$ | SD | | $X^2$ |
|---|---|---|---|---|---|---|
| | -1 | +1 | | -2 | +2 | |
| 1......... | 45 | 12 | 19.105*** | 5 | 6 | |
| 2......... | 10 | 24 | 5.765* | 0 | 2 | |
| 3......... | 13 | 21 | | 1 | 7 | |
| 4......... | 16 | 8 | | 0 | 2 | |
| 5......... | 6 | 29 | 13.828*** | 0 | 3 | |
| 6......... | 22 | 6 | 8.036** | 2 | 1 | |
| 7......... | 13 | 13 | | 0 | 1 | |
| 8......... | 11 | 12 | | 0 | 4 | |
| 9......... | 17 | 11 | | 0 | 1 | |
| 10......... | 12 | 11 | | 0 | 2 | |
| 11......... | 10 | 7 | | 0 | 1 | |
| 12......... | 16 | 11 | | 0 | 0 | |
| 13......... | 13 | 11 | | 0 | 2 | |
| 14......... | 6 | 26 | 12.500*** | 0 | 6 | |
| Total | 210 | 202 | | 8 | 38 | 19.565*** |

For 1 df: $^{*}P(X^2 > 3.841) \doteq .05$
$^{**}P(X^2 > 6.635) \doteq .01$
$^{***}P(X^2 > 10.827) = .001$

## TABLE XII

### PICTURES DEVIATING ±1 SD FROM INDIVIDUAL MEAN WORD PRODUCTIVITY FOR COMPARATIVE GROUPS

| | Community | | | | Sex | | | | Grade | | | | Intelligence | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rural | | Urban | | Boy | | Girl | | First | | Sixth | | High* | | Low* | |
| Picture | -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 |
| 1......... | 22** | 5** | 23** | 7** | 23 | 6 | 22 | 6 | 19 | 8 | 26 | 4 | 9 | 0 | 9 | 2 |
| 2......... | 5 | 12 | 5 | 12 | 4 | 13 | 6 | 11 | 5 | 16 | 5 | 8 | 1 | 5 | 3 | 4 |
| 3......... | 7 | 8 | 6 | 13 | 5 | 11 | 8 | 10 | 8 | 9 | 5 | 12 | 2 | 4 | 4 | 1 |
| 4......... | 9 | 6 | 7 | 2 | 7 | 7 | 9 | 1 | 10 | 4 | 6 | 4 | 6 | 2 | 2 | 0 |
| 5......... | 4 | 16 | 2 | 13 | 3 | 12 | 3 | 17 | 5 | 12 | 1 | 17 | 1 | 2 | 0 | 7 |
| 6......... | 8 | 3 | 14 | 3 | 11 | 3 | 11 | 3 | 7 | 5 | 15 | 1 | 7 | 0 | 4 | 4 |
| 7......... | 8 | 9 | 5 | 4 | 4 | 7 | 9 | 6 | 6 | 10 | 7 | 3 | 2 | 2 | 1 | 1 |
| 8......... | 8 | 8 | 3 | 4 | 4 | 9 | 7 | 3 | 6 | 4 | 5 | 8 | 1 | 3 | 3 | 2 |
| 9......... | 6 | 5 | 11 | 6 | 11 | 3 | 6 | 8 | 9 | 5 | 8 | 6 | 4 | 3 | 5 | 0 |
| 10......... | 4 | 8 | 8 | 3 | 8 | 4 | 4 | 7 | 6 | 4 | 6 | 7 | 0 | 5 | 1 | 2 |
| 11......... | 4 | 3 | 6 | 4 | 3 | 1 | 7 | 6 | 6 | 4 | 4 | 3 | 1 | 4 | 2 | 1 |
| 12......... | 9 | 5 | 7 | 6 | 9 | 5 | 7 | 6 | 10 | 5 | 6 | 6 | 3 | 2 | 3 | 1 |
| 13......... | 6 | 5 | 7 | 6 | 8 | 5 | 5 | 6 | 6 | 6 | 7 | 5 | 1 | 2 | 1 | 1 |
| 14......... | 5 | 10 | 1 | 16 | 3 | 11 | 3 | 15 | 6 | 10 | 0 | 16 | 0 | 7 | 0 | 4 |
| Total | 105 | 103 | 105 | 99 | 103 | 97 | 107 | 105 | 109 | 102 | 101 | 100 | 38 | 41 | 38 | 30 |

*N = 16 instead of 44 as for other groups

**$X^2$ (Distribution) for each picture in each group: P = not significant.

TABLE XIII

PICTURES DEVIATING ±2 SD FROM INDIVIDUAL MEAN WORD PRODUCTIVITY FOR COMPARATIVE GROUPS

| | Community | | | | Sex | | | | Grade | | | | Intelligence | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rural | | Urban | | Boy | | Girl | | First | | Sixth | | High* | | Low* | |
| Picture | -2 | +2 | -2 | +2 | -2 | +2 | -2 | +2 | -2 | +2 | -2 | +2 | -2 | +2 | -2 | +2 |
| 1......... | 2** | 2** | 3** | 4** | 2 | 2 | 3 | 4 | 2 | 4 | 3 | 2 | 0 | 0 | 1 | 2 |
| 2......... | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 3......... | 1 | 2 | 0 | 5 | 1 | 5 | 0 | 2 | 1 | 3 | 0 | 4 | 0 | 2 | 1 | 0 |
| 4......... | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 5......... | 0 | 2 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 2 |
| 6......... | 0 | 1 | 2 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 1 |
| 7......... | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8......... | 0 | 3 | 0 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 0 | 2 |
| 9......... | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 10......... | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| 11......... | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 12......... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13......... | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| 14......... | 0 | 2 | 0 | 4 | 0 | 2 | 0 | 4 | 0 | 3 | 0 | 3 | 0 | 1 | 0 | 0 |
| Total | 3 | 19 | 5 | 19 | 4 | 22 | 4 | 16 | 3 | 16 | 5 | 22 | 0 | 8 | 3 | 8 |

*N = 16 instead of 44 as for other groups.

**$X^2$ (Distribution) for each picture in each group: P = not significant.

## TABLE XIV

## PICTURES DEVIATING ±1 SD FROM INDIVIDUAL MEAN WORD PRODUCTIVITY OF RURAL AND URBAN GROUPS ANALYZED BY LIFE-SITUATION PATTERNS

| Life-Situation Picture | | | | Rural | | | Urban | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Pattern | Group | No. | Name | -1 SD | +1 SD | Total | -1 SD | +1 SD | Total | Probability |
| Community | Rural | 13 | Village | 6 | 5 | 11 | 7 | 6 | 13 | Not Significant |
| | | 7 | Farm | 8 | 9 | 17 | 5 | 4 | 9 | |
| | | | Total | 14 | 14 | 28 | 12 | 10 | 22 | |
| | Urban | 10 | City | 4 | 8 | 12 | 8 | 3 | 11 | |
| | | 4 | Factory | 9 | 6 | 15 | 7 | 2 | 9 | |
| | | | Total | 13 | 14 | 27 | 15 | 5 | 20 | |
| Social Status | High | 5 | Mansion | 4 | 16 | 20 | 2 | 13 | 15 | Not Significant |
| | | 11 | Resort | 4 | 3 | 7 | 6 | 4 | 10 | |
| | | | Total | 8 | 19 | 27 | 8 | 17 | 25 | |
| | Low | 2 | Hovel | 5 | 12 | 17 | 5 | 12 | 17 | |
| | | 8 | Swim. Hole | 8 | 8 | 16 | 3 | 4 | 7 | |
| | | | Total | 13 | 20 | 33 | 8 | 16 | 24 | |
| Age | Child | 1 | Bedroom | 22 | 5 | 27 | 23 | 7 | 30 | Not Significant |
| | | 6 | Church | 8 | 3 | 11 | 14 | 3 | 17 | |
| | | 12 | Schoolroom | 9 | 5 | 14 | 7 | 6 | 13 | |
| | | | Total | 39 | 13 | 52 | 44 | 16 | 60 | |
| | Adult | 3 | Dam | 7 | 8 | 15 | 6 | 13 | 19 | |
| | | 9 | Capitol | 6 | 5 | 11 | 11 | 6 | 17 | |
| | | 14 | Dock | 5 | 10 | 15 | 1 | 16 | 17 | |
| | | | Total | 18 | 23 | 41 | 18 | 35 | 53 | |

## TABLE XV

### PICTURES DEVIATING ±1 SD FROM INDIVIDUAL MEAN WORD PRODUCTIVITY OF BOYS AND GIRLS ANALYZED BY LIFE-SITUATION PATTERNS

| Life-Situation Picture | | | | Boys | | | Girls | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Pattern | Group | No. | Name | -1 SD | +1 SD | Total | -1 SD | +1 SD | Total | Probability |
| Community | Rural | 13 | Village | 8 | 5 | 13 | 5 | 6 | 11 | Not Significant |
| | | 7 | Farm | 4 | 7 | 11 | 9 | 6 | 15 | |
| | | | Total | 12 | 12 | 24 | 14 | 12 | 26 | |
| | Urban | 10 | City | 8 | 4 | 12 | 4 | 7 | 11 | |
| | | 4 | Factory | 7 | 7 | 14 | 9 | 1 | 10 | |
| | | | Total | 15 | 11 | 26 | 13 | 8 | 21 | |
| Social Status | High | 5 | Mansion | 3 | 12 | 15 | 3 | 17 | 20 | *+1 SD Distribution:<br>$X^2 = 4.502$<br>df = 1<br>$P < .05$ |
| | | 11 | Resort | 3 | 1 | 4 | 7 | 6 | 13 | |
| | | | Total | 6 | 13* | 19 | 10 | 23* | 33 | |
| | Low | 2 | Hovel | 4 | 13 | 17 | 6 | 11 | 17 | |
| | | 8 | Swim. Hole | 4 | 9 | 13 | 7 | 3 | 10 | |
| | | | Total | 8 | 22* | 30 | 13 | 14* | 27 | |
| Age | Child | 1 | Bedroom | 23 | 6 | 29 | 22 | 6 | 28 | Not Significant |
| | | 6 | Church | 11 | 3 | 14 | 11 | 3 | 14 | |
| | | 12 | Schoolroom | 9 | 5 | 14 | 7 | 6 | 13 | |
| | | | Total | 43 | 14 | 57 | 40 | 15 | 55 | |
| | Adult | 3 | Dam | 5 | 11 | 16 | 8 | 10 | 18 | |
| | | 9 | Capitol | 11 | 3 | 14 | 6 | 8 | 14 | |
| | | 14 | Dock | 3 | 11 | 14 | 3 | 15 | 18 | |
| | | | Total | 19 | 25 | 44 | 17 | 33 | 50 | |

## TABLE XVI

## PICTURES DEVIATING $\underline{+}1$ SD FROM INDIVIDUAL MEAN WORD PRODUCTIVITY OF FIRST AND SIXTH GRADE GROUPS ANALYZED BY LIFE-SITUATION PATTERNS

| Life-Situation Picture | | | | 1st Grade | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Pattern | Group | No. | Name | -1 SD | +1 SD | Total | -1 SD | +1 SD | Total | Probability |
| Community | Rural | 13 | Village | 6 | 6 | 12 | 7 | 5 | 12 | Not Significant |
| | | 7 | Farm | 6 | 10 | 16 | 7 | 3 | 10 | |
| | | | Total | 12 | 16 | 28 | 14 | 8 | 22 | |
| | Urban | 10 | City | 6 | 4 | 10 | 6 | 7 | 13 | |
| | | 4 | Factory | 10 | 4 | 14 | 6 | 4 | 10 | |
| | | | Total | 16 | 8 | 24 | 12 | 11 | 23 | |
| Social Status | High | 5 | Mansion | 5 | 12 | 17 | 1 | 17 | 18 | Not Significant |
| | | 11 | Resort | 6 | 4 | 10 | 4 | 3 | 7 | |
| | | | Total | 11 | 16 | 27 | 5 | 20 | 25 | |
| | Low | 2 | Hovel | 5 | 16 | 21 | 5 | 8 | 13 | |
| | | 8 | Swim. Hole | 6 | 4 | 10 | 5 | 8 | 13 | |
| | | | Total | 11 | 20 | 31 | 10 | 16 | 26 | |
| Age | Child | 1 | Bedroom | 19 | 8 | 27 | 26 | 4 | 30 | *-1 SD Distribution:<br>$X^2 = 4.308$<br>df = 1<br>$P < .05$ |
| | | 6 | Church | 7 | 5 | 12 | 15 | 1 | 16 | |
| | | 12 | Schoolroom | 10 | 5 | 15 | 6 | 6 | 12 | |
| | | | Total | 36* | 18 | 54 | 47* | 11 | 58 | |
| | Adult | 3 | Dam | 8 | 9 | 17 | 5 | 12 | 17 | |
| | | 9 | Capitol | 9 | 5 | 14 | 8 | 6 | 14 | |
| | | 14 | Dock | 6 | 10 | 16 | 0 | 16 | 16 | |
| | | | Total | 23* | 24 | 47 | 13* | 34 | 47 | |

## TABLE XVII

### PICTURES DEVIATING ±1 SD FROM INDIVIDUAL MEAN WORD PRODUCTIVITY OF HIGH AND LOW INTELLIGENCE GROUPS ANALYZED BY LIFE-SITUATION PATTERNS

| Life-Situation Picture | | | | High Intelligence | | | Low Intelligence | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Pattern | Group | No. | Name | -1 SD | +1 SD | Total | -1 SD | +1 SD | Total | Probability |
| Community | Rural | 13 | Village | 1 | 2 | 3 | 1 | 1 | 2 | Not Significant |
| | | 7 | Farm | 2 | 2 | 4 | 1 | 1 | 2 | |
| | | | Total | 3 | 4 | 7 | 2 | 2 | 4 | |
| | Urban | 10 | City | 0 | 5 | 5 | 1 | 2 | 3 | |
| | | 4 | Factory | 6 | 2 | 8 | 2 | 0 | 2 | |
| | | | Total | 6 | 7 | 13 | 3 | 2 | 5 | |
| Social Status | High | 5 | Mansion | 1 | 2 | 3 | 0 | 7 | 7 | Not Significant |
| | | 11 | Resort | 1 | 4 | 5 | 2 | 1 | 3 | |
| | | | Total | 2 | 6 | 8 | 2 | 8 | 10 | |
| | Low | 2 | Hovel | 1 | 5 | 6 | 3 | 4 | 7 | |
| | | 8 | Swim. Hole | 1 | 3 | 4 | 3 | 2 | 5 | |
| | | | Total | 2 | 8 | 10 | 6 | 6 | 12 | |
| Age | Child | 1 | Bedroom | 9 | 0 | 9 | 9 | 2 | 11 | * +1 SD Distribution:<br>$X^2 = 4.667$<br>df = 1<br>$P < .05$<br>(Yates' correction applied) |
| | | 6 | Church | 7 | 0 | 7 | 4 | 4 | 8 | |
| | | 12 | Schoolroom | 3 | 2 | 5 | 3 | 1 | 4 | |
| | | | Total | 19 | 2* | 21 | 16 | 7* | 23 | |
| | Adult | 3 | Dam | 2 | 4 | 6 | 4 | 1 | 5 | |
| | | 9 | Capitol | 4 | 3 | 7 | 5 | 0 | 5 | |
| | | 14 | Dock | 0 | 7 | 7 | 0 | 4 | 4 | |
| | | | Total | 6 | 14* | 20 | 9 | 5* | 14 | |

view that develops with age. In short, first-grade children haven't yet broadened their interests to include more remote situations, whereas sixth-grade children no longer find great stimulation in a consideration of more immediate experiences. The equating of community and sex in the grade groups would discount the possibility of differences appearing in the community and social-status patterns, which prediction was borne out by the findings.

The responsiveness of bright children to adult or remote experiences and that of less intelligent pupils to child or immediate experiences can likewise be interpreted from the standpoint of the relationship between mental development and the growth of interests and concerns. These groups being equated in community, sex, and grade variables, there would be no basis for predicting differences with respect to the other patterns of life-situations, and none were found.

### Is Word Productivity Related to Picture Preference?

Several hypotheses can be advanced regarding the verbal output of children on pictures which they like best or like least. One would be that children have a great deal to say about what pleases them, and very little about situations which they dislike. It could be, too, that children take for granted the desirability of their preferences, but find it necessary to substantiate their rejections with extended statements. On the other hand, it may be that it requires as much verbal activity to defend choices as it does to substantiate rejections.

Data for the examination of this question were provided in Part III of the Social Perception Interview. The subject was asked to designate "the picture you would like to keep for your very own" and to tell why. He was given a copy of this picture and requested to select a second, and then a third in similar fashion. Finally, he was asked to point out "the picture you wouldn't like to keep," and to tell why.

Word Productivity for the first, second, third, and least preferred and the 10 pictures which were not mentioned is given for the total group in Table XVIII, and for the comparative groups in Table XIX. The skewness of each distribution toward the lower end of the scale in the former table corroborates the finding of the preceding section that deviant responses are more likely to be in the direction of greater productivity rather than less.

Comparisons in mean Word Productivity for the various designations of likes and dislikes as well as for the combined preferences and non-selected pictures are given in Table XX.. The only instances of statistically significant differences occur in the comparison of the Intelligence groups. The difference between the productivity of the third choice and least preferred picture is significant at the 5 percent level, and that between the second and least preferred is significant at the 1 percent level of confidence. In both cases, the productivity of the rejected picture is higher than that of the preference pictures.

The evidence indicates that, generally speaking, word output in the free-association responses (Part I) is no predictor of picture preferences in Part III of the Social-Perception Interview. It would seem that whatever the causative factors operating in Part I, they are not identical to those influencing picture-selection behavior in Part III. It may be that telling a story about a situation draws upon the individual's background of experience whereas, in the selection of pictures, values and interests are brought into focus. The difference is between what the subject knows and is familiar with (Part I) and what he would like to have or hopes to be (Part III). Whether background or aspiration is involved, the same number of words is required for the expression of one as for the other.

Although this explanation also applies to the Low Intelligence group, it does not explain why exceptions should occur in the relationship between second and third choices with that of the least preferred picture. One explanation is that less intelligent children have more limited goals than bright children, and that once having selected the "best" picture they were less able or less motivated to discriminate among the remaining pictures for second and third choices.

### Conclusions

A number of important conclusions may be reached regarding the free-association responses to the Life-Situation Picture Series. First, Word Productivity is an index of behavior. Pupils did not respond to the pictures in a random fashion; neither were word responses a reflection of certain innate pupil abilities or characteristics of the instrument. Second, there is a relationship between children's social background and their Word Productivity on specific life-situation pictures or to patterns of these situations. These differences were related to intelligence, grade, and sex in descending order, there being none associated with the rural and urban populations sampled in this study. Third, the behaviors elicited in the free-association and picture-preference phases of the Social-Perception Interview are different. Whereas an analysis of Word Productivity in terms of children's social background did reveal significant relationships, this was not the case for the analysis in terms of picture choices. This is in keeping with the design for the Social-Perception Interview which was planned to bring about increasing involvement on the part of subjects.

## TABLE XVIII

## WORD PRODUCTIVITY OF PICTURE PREFERENCES FOR TOTAL GROUP

| | Picture Preference | | | | Total | |
|---|---|---|---|---|---|---|
| Words | 1st | 2nd | 3rd | Least | Preferred (1st, 2nd, 3rd) | Non-Selected (N=10) |
| 210-219..... | | | | | | 1 |
| 200-209..... | | | | | | 0 |
| 190-199..... | | | | | | 1 |
| 180-189..... | 1 | | | | 1 | 0 |
| 170-179..... | 0 | | | | 0 | 0 |
| 160-169..... | 0 | | | | 0 | 4 |
| 150-159..... | 0 | 1 | 1 | | 2 | 1 |
| 140-149..... | 0 | 0 | 0 | 1 | 0 | 2 |
| 130-139..... | 0 | 0 | 0 | 1 | 0 | 4 |
| 120-129..... | 0 | 1 | 0 | 0 | 1 | 8 |
| 110-119..... | 0 | 1 | 3 | 1 | 4 | 11 |
| 100-109..... | 2 | 5 | 3 | 2 | 10 | 10 |
| 90- 99..... | 2 | 0 | 3 | 2 | 5 | 20 |
| 80- 89..... | 2 | 1 | 1 | 1 | 4 | 25 |
| 70- 79..... | 5 | 1 | 2 | 6 | 8 | 26 |
| 60- 69..... | 6 | 5 | 4 | 2 | 15 | 49 |
| 50- 59..... | 7 | 9 | 4 | 8 | 20 | 66 |
| 40- 49..... | 6 | 12 | 12 | 12 | 30 | 100 |
| 30- 39..... | 12 | 14 | 11 | 19 | 37 | 137 |
| 20- 29..... | 18 | 15 | 20 | 10 | 53 | 183 |
| 10- 19..... | 20 | 14 | 15 | 16 | 49 | 169 |
| 0- 9..... | 7 | 9 | 8 | 6 | 24 | 65 |
| N | 88 | 88 | 87* | 87* | 263 | 882 |
| Mean | 38.41 | 40.46 | 40.29 | 41.67 | 39.72 | 39.56 |
| SD | 29.63 | 30.21 | 31.25 | 29.08 | 30.21 | 29.57 |
| $SE_M$ | 3.16 | 3.22 | 3.35 | 3.12 | | |

*Preference not indicated by 1st grade rural girl and by 1st grade urban girl respectively.

## TABLE XIX

## WORD PRODUCTIVITY OF PICTURE PREFERENCES FOR COMPARATIVE GROUPS

| Comparison | Group | Statistic | Picture Preference |  |  |  | Total |  |
|---|---|---|---|---|---|---|---|---|
|  |  |  | 1st | 2nd | 3rd | Least | Preferred (1, 2, 3) | Non-Selected |
| Community | Rural | N | 44 | 44 | 43 | 44 | 131 | 441 |
|  |  | Mean | 37.73 | 39.77 | 35.93 | 40.45 | 37.82 | 38.38 |
|  |  | SD | 34.19 | 34.67 | 31.68 | 30.69 | 33.21 | 31.53 |
|  |  | SE$_M$ | 5.16 | 5.23 | 4.83 | 4.63 | * | * |
|  | Urban | N | 44 | 44 | 44 | 43 | 132 | 441 |
|  |  | Mean | 39.09 | 43.41 | 44.55 | 40.58 | 41.59 | 40.74 |
|  |  | SD | 24.62 | 24.97 | 30.57 | 28.23 | 26.77 | 27.41 |
|  |  | SE$_M$ | 3.71 | 3.72 | 4.61 | 4.30 | * | * |
| Sex | Boy | N | 44 | 44 | 44 | 44 | 132 | 440 |
|  |  | Mean | 41.14 | 40.91 | 40.23 | 39.77 | 40.76 | 39.48 |
|  |  | SD | 32.15 | 28.47 | 33.38 | 28.08 | 31.05 | 30.06 |
|  |  | SE$_M$ | 4.85 | 4.29 | 5.03 | 4.23 | * | * |
|  | Girl | N | 44 | 44 | 43 | 43 | 131 | 442 |
|  |  | Mean | 35.68 | 40.00 | 40.35 | 43.61 | 38.66 | 39.64 |
|  |  | SD | 26.98 | 32.17 | 29.31 | 30.28 | 29.30 | 29.07 |
|  |  | SE$_M$ | 4.07 | 4.85 | 4.47 | 4.62 | * | * |
| Grade | First | N | 44 | 44 | 43 | 43 | 131 | 442 |
|  |  | Mean | 34.77 | 34.77 | 38.72 | 41.05 | 36.07 | 36.15 |
|  |  | SD | 26.53 | 27.32 | 35.32 | 34.24 | 29.74 | 30.06 |
|  |  | SE$_M$ | 4.00 | 4.12 | 5.39 | 5.22 | * | * |
|  | Sixth | N | 44 | 44 | 44 | 44 | 132 | 440 |
|  |  | Mean | 42.05 | 46.14 | 41.82 | 42.27 | 43.33 | 42.98 |
|  |  | SD | 32.10 | 32.15 | 27.00 | 23.36 | 30.23 | 28.66 |
|  |  | SE$_M$ | 4.48 | 4.85 | 4.07 | 3.52 | * | * |
| Intelligence | High** | N | 16 | 16 | 16 | 16 | 48 | 160 |
|  |  | Mean | 51.25 | 55.13 | 55.44 | 48.56 | 54.38 | 51.38 |
|  |  | SD | 29.81 | 38.48 | 34.52 | 31.17 | 33.61 | 30.50 |
|  |  | SE$_M$ | 7.45 | 9.62 | 8.63 | 7.79 | * | * |
|  | Low** | N | 16 | 16 | 16 | 15 | 48 | 161 |
|  |  | Mean | 35.75 | 25.38 | 29.06 | 35.53 | 30.42 | 31.96 |
|  |  | SD | 43.01 | 24.71 | 26.89 | 28.82 | 32.02 | 33.52 |
|  |  | SE$_M$ | 10.76 | 6.18 | 6.72 | 7.44 | * | * |

* Because of the difference in N between Total Preferred and Total Non-selected pictures, SE$_M$ was not computed separately, but combined in formula:

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\left(\frac{N_1\sigma_1^2 + N_2\sigma_2^2}{N_1 + N_2 - 2}\right)\left(\frac{N_1 + N_2}{N_1 N_2}\right)}$$

**N = 16 subjects instead of 44 as for other groups.

## TABLE XX

## COMPARISON OF MEAN WORD PRODUCTIVITY OF PICTURE PREFERENCES AND NON-SELECTED PICTURES

| Comparison | Group | 1st, 2nd, 3rd, and Least Preferred Pictures | | | | | | Total Preferred-Non-Selected* |
|---|---|---|---|---|---|---|---|---|
| | | 1st-2nd | 1st-3rd | 1st-Lst | 2nd-3rd | 2nd-Lst | 3rd-Lst | |
| Community | Rural | t = .279 | .254 | .394 | .540 | .098 | .676 | .174 |
| | Urban | .821 | .922 | .262 | .192 | .497 | .629 | .315 |
| Sex | Boy | .035 | .130 | .212 | .104 | .190 | .069 | .425 |
| | Girl | .682 | .772 | 1.287 | .053 | .538 | .507 | .336 |
| Grade | First | .000 | .588 | .954 | .582 | .943 | .310 | .028 |
| | Sixth | .597 | .036 | .038 | .682 | .645 | .085 | .123 |
| Intelligence | High** | .318 | .367 | .249 | .024 | .530 | .591 | .581 |
| | Low** | .836 | .527 | .063 | .404 | $4.009^1$ | $2.459^2$ | .281 |
| Total (N = 88) | | .454 | .408 | .734 | .036 | .270 | .302 | .065 |

* Preferred = 1st, 2nd, 3rd choices; non-selected = 10 pictures. For 30 df: $P(t > 2.750) = .01^1$
$P(t > 2.042) = .05^2$

**N = 16 instead of 44 as for other groups.

In short, Word Productivity provides one source of validation for the instrument. Gross as this measure of human behavior may be, it reveals certain significant relationships which, with few exceptions, can be explained in terms of what is known about children. Equally important, an analysis of word count, with the exceptions noted, has failed to produce relationships where none could reasonably be expected.

A number of generalizations also pertain to the nature of Word Productivity itself. From the outset, it was established that there is a difference between word knowledge, as commonly determined by vocabulary tests, and the quantitative aspect of verbal behavior in spontaneous oral expression. That an individual has a better grasp of word meanings is no indicator of how many words he will use in expressing himself. Also, it is clear, that broad generalizations about superiority in language facility are over-simplifications. No one group of children showed consistently greater response or lack of response to the Life-Situation Picture Series. Word Productivity depends on *what* is being talked about. Studies of verbal facility will produce different results depending upon whether subjects are free to select topics for discussion or are required to deal with prescribed situations.

Two cautions must be pointed out for the interpretation of the above findings. First, there is considerable indeterminancy regarding Word Productivity when analyzed from the standpoint of deviations from each individual's mean word response. Because of limited sampling groups many frequencies were zero or too small to be treated statistically. To clarify these indeterminate areas, quite obviously, will require larger sampling procedures.

The fact that none of the analyses revealed rural-urban differences raises the question of the applicability of these findings. Although the county being sampled has only five communities, none of which are over 15,000 population, and all the subjects were farm children, does this constitute a rural sample? In similar terms, although the urban community is located in the second largest urban region in this country, do these children represent urban children? Could it be that children living in more isolated rural regions would differ from those living in a metropolitan community of several million population? It may be that to compare rural and urban children *in general* is another instance of over-simplification, and is one explanation for the lack of agreement found in studies of these culture groups.

## FOOTNOTES

* For assistance with statistical computations, the writer wishes to acknowledge his indebtedness to the following research assistants: Irvin J. Lehmann, Sobhi T. Geraissa, Donald W. Hinkkanen, and Mohamed K. Hindy, as well as to R. James Evey, Project Supervisor of the Numerical Analysis Laboratory, University of Wisconsin.

1. Frank J. Estvan and Elizabeth W. Estvan. *The Child's World: His Social Perception* (New York: G. P. Putnam's Sons, 1959).

2. Frank J. Estvan. "Studies in Social Perception: Methodology," *Journal of Genetic Psychology*, XCII (June 1958), pp. 215-46.

3. 1 = First grade 6 = Sixth grade
R = Rural U = Urban
B = Boy G = Girl

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 1, September 1959)

# THE RELATIONSHIP BETWEEN HANDWRITING PRESSURE AND LEGIBILITY OF HANDWRITING IN CHILDREN AND ADOLESCENTS*

THEODORE L. HARRIS and G. LAWRENCE RARICK
University of Wisconsin

## Statement of the Problem

THE PURPOSE of this investigation was to examine the relationship between handwriting legibility and point pressure in the handwriting of boys and girls in grades 4, 6, and 10. Previous research at the college level had indicated that high legibility tends to be accompanied by relatively low variability in point force, while low legibility tends to be accompanied by relatively high variability in point force. This study was specifically designed to examine the relationship between these two factors at younger age levels.

## The Background of the Investigation

The present research is one of several studies which are being conducted at the University of Wisconsin by a Committee for Research in Handwriting.[1] The Committee works as a unit in planning the research program. In addition, specific aspects of research are delegated to individual members of the group. Aspects of handwriting currently being studied by the Committee include studies of the characteristics of the handwriting product in terms of legibility and its associated factors, the characteristics of effective handwriting instruments, the factors influencing choice of handwriting instruments, motor components involved in the handwriting process, and perceptual factors involved in the production and evaluation of handwriting.

The present research is a direct extension of an investigation of the relationship between handwriting pressure and legibility at the college level, originally reported in 1955 as a bulletin of the School of Education of the University of Wisconsin, and subsequently reprinted in the Journal of Experimental Education.[2] This study should be consulted for a discussion of previous research concerning handwriting pressure and for a presentation of the technical details of the instrumentation used and the type of records obtained.

The experimental equipment employed in the present investigation was the same as that used in the college study. Essentially, the equipment included a specially constructed writing table with a metal platen mounted flush with an adjustable section of the table top (see Figure 1). The platen rested on a strain gauge system which made possible the detection of minute changes in point pressure during writing. The signal from the strain gauge system was amplified by a vacuum tube amplifier and photographically recorded by a Hathaway recording oscillograph.

It became clear at an early stage in the analysis of the oscillographic records that two or more records having the same average word or sentence force might differ markedly in the constancy with which the force was applied to the page in writing. It was felt, further, that if the amount of variation in force could be measured in such a way as to make comparisons between records of varying speeds and levels of average force, a new way might be found to explore the relationship between handwriting force and legibility.

---

* Funds for the support of this study were provided in part by a grant from the Parker Pen Company, Janesville, Wisconsin, to the University of Wisconsin, and in part by a research grant from the Graduate School of the University of Wisconsin. Valuable assistance in this research was furnished by Mary Haberkorn Juaire, Research Assistant, and Henry K. Kaplan, then a graduate student in the Department of Education.

[1] The Committee is composed of the following staff members: Paul W. Eberman, Professor of Education; Theodore L. Harris, Professor of Education; Virgil E. Herrick, Chairman, Professor of Education; and G. Lawrence Rarick, Professor of Physical Education.

[2] Theodore L. Harris and G. Lawrence Rarick. "The Problem of Pressure in Handwriting," Journal of Experimental Education, XXVI (December 1957), pp. 151-78.

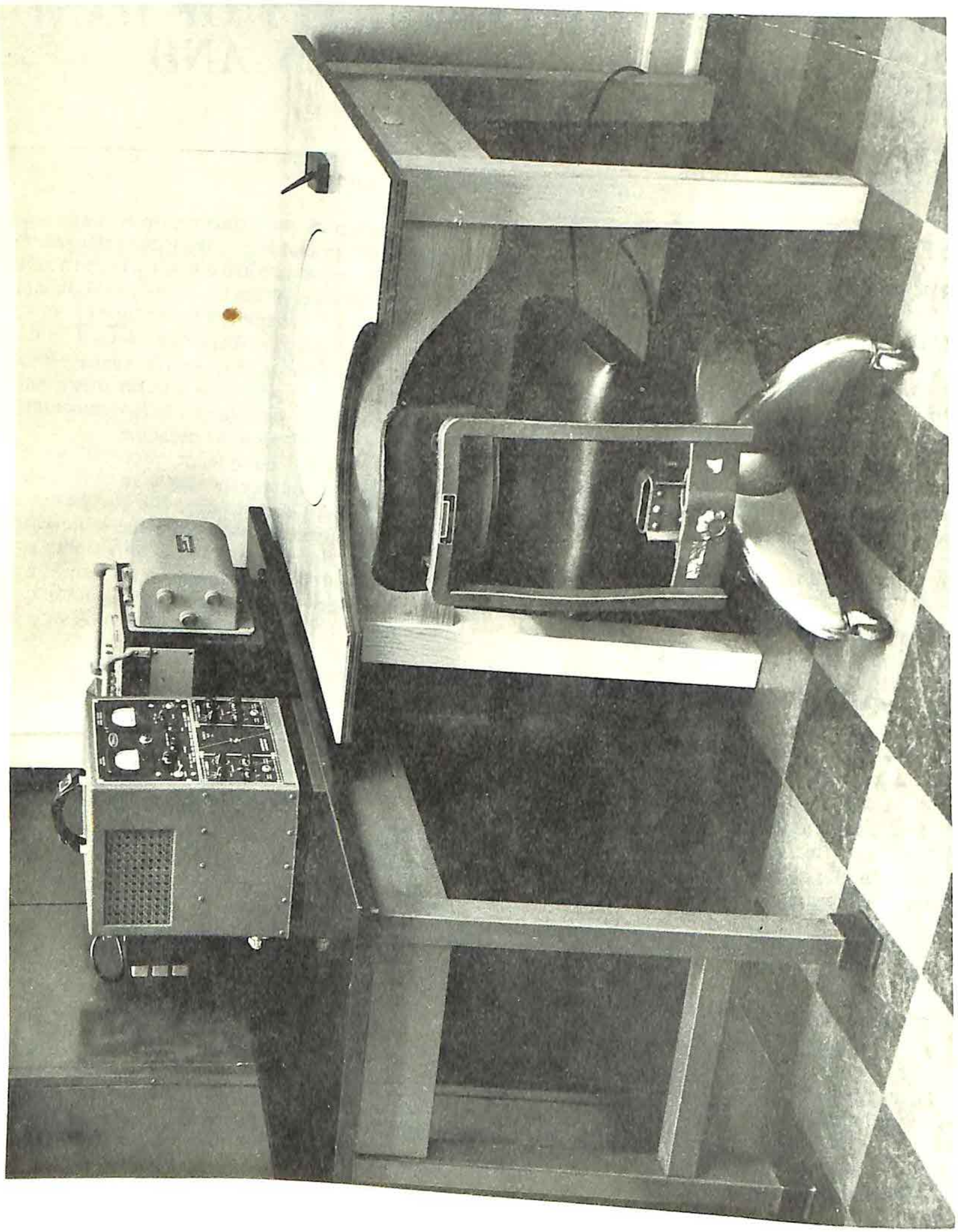


FIGURE 1

WRITING TABLE WITH METAL PLATEN AND ADJUSTABLE SECTION OF TABLE TOP

A step toward the solution of this problem was made in the development of the force variation ratio, the derivation of which is described in the college study.[3] This technique of analysis makes it possible to measure the cumulative amount of change in the application of force for each word, and to derive a ratio between the sum of the deviations in force from the base line of the oscillographic record and the time taken to produce the record. The formula for the force variation ratio is:

$$\text{FVR} = \frac{\text{Change in mm. -2 (average height of curve)}}{\text{Length of base line (time)}}$$

This ratio was used as a major technique in the analysis of the oscillographic pressure records in the college study.

Findings of the College Study— Two types of findings in the original study will be mentioned briefly here, one relating to data gathered concerning force variation ratio, and the other to the relationship found between this ratio and legibility.

One of the concerns of the original study was to investigate the significance and stability of the force variation ratio. Accordingly, a test of fine motor coordination was devised; namely, to draw four one-inch horizontal lines as steadily as possible. The point force was recorded and analyzed. The stability of this test was indicated by the relatively high correlation of .88 between the sum of the force variation ratio for the first and fourth and for the second and third samples, respectively, for 19 college subjects. The stability of the pressure pattern in tasks which are similar in nature is shown by the fact that a correlation of .81 was secured between the force variation ratios for the short line test and those of the normal handwriting samples for the same group of subjects. This suggests that the force variation ratio provides a measure of fine motor control in handwriting.

The relationships among factors studied in the college investigation as indicated by correlation analysis are reported in Table I. Note that the most substantial correlations involve the force variation ratio. Average pressure, on the other hand, shows little relationship to any other variable except variation in force. These findings markedly influenced the decision to study further the relationship between the force variation ratio and legibility in a population of children and adolescents.

Methodology of the Current Study

Population and Design—The study was designed so that an analysis of variance technique might be used in the treatment of the data. The subjects were 144 boys and girls randomly drawn from the schools of two small urban communities in southern Wisconsin in such a way that 12 boys and 12 girls were secured at each of three grade levels—grades 4, 6, and 10—in each community. It should be pointed out that the population was further limited to those children who were right-handed, who had had continuous experience in the school systems concerned, and who were within the normal range of age expectancy for their grade. Mean chronological ages of the subjects by sex and grade are given in Table II. In those cases in which more than one school in a given community was involved, cases were drawn on a proportional basis according to the number of children per grade per school. It was felt that this procedure in the selection of the experimental population would make possible a more systematic approach to the study of the relationship between the force variation ratio and legibility as it might be affected by the factors of sex and community.

The Experimental Task—The basic data for this study were gathered during the fall and winter of the 1955-56 academic year. Since a relatively large population of children was included in the study it seemed best to move the experimental equipment to the two communities in which the data were to be gathered. The cooperating schools provided a room for the experimental equipment and also aided in securing an experienced teacher to assist in helping the children adjust to the experimental situation.

Each child was tested individually, all measures being secured on a particular child during one testing period. Upon entering the testing room the child was seated at the handwriting table and necessary adjustments in chair height and writing angle were made. Each child was asked to write his name on a card. The child was then shown a card with the typewritten sentence, "The quick brown fox jumps over the lazy dog." He was asked to study the sentence until he could write it from memory. Each pupil was given an opportunity to write the standard sentence several times until he learned to spell each word and to reproduce the entire sentence correctly. Next, a fresh card was placed over the platen in preparation for recording point pressure in writing the standard sentence.

After determining that the pupil could repeat the standard sentence from memory, he was then asked to write the sentence on the card. Care was taken to instruct each child to write in his usual way. At this time his handwriting pressure was recorded. After completing the reproduction of the standard handwriting sample, a second card was placed over the platen (see Figure 2). The subject was asked to complete the figure with either curved or straight lines as indicated on the

---

[3] Ibid., p. 162.

## TABLE I

## RANK-ORDER CORRELATIONS BETWEEN LEGIBILITY TIME, AVERAGE PRESSURE, AND VARIATION IN FORCE FOR 57 NORMAL, SLOW, AND FAST ADULT SAMPLES

| Variables | Rho |
|---|---|
| Variation in force[a] and pressure[b] | .69 |
| Variation in force and time[c] | .67 |
| Variation in force and legibility[d] | .59 |
| Legibility and time | .54 |
| Pressure and time | .22 |
| Pressure and legibility | .01 |

[a] Ranked from low to high
[b] Ranked from low to high
[c] Ranked from slowest to fastest
[d] Ranked from best to poorest

## TABLE II

## MEAN CHRONOLOGICAL AGES OF SUBJECTS BY GRADE AND SEX

| | Boys | | | Girls | | |
|---|---|---|---|---|---|---|
| Grade | N | Mean (Months) | S. D. | N | Mean (Months) | S. D. |
| Grade 4 | 24 | 116.38 | 4.80 | 24 | 115.75 | 4.38 |
| Grade 6 | 24 | 143.88 | 9.81 | 24 | 143.50 | 11.72 |
| Grade 10 | 24 | 185.13 | 7.96 | 24 | 186.08 | 5.63 |

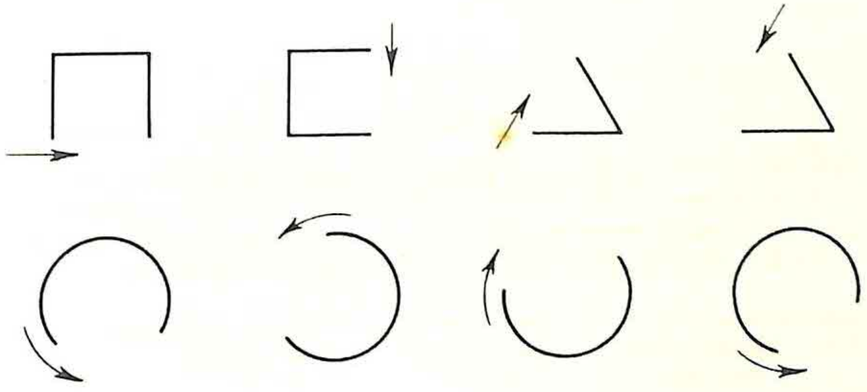


FIGURE 2: FORMS TO BE COMPLETED IN THE TEST OF FINE MOTOR CONTROL

card. This test was designed to measure the pupil's finer motor control in a situation similar to that of handwriting.

Methods of Analysis of Data—In the present study, force variation ratios were calculated by the techniques used in the college study. However, the calculation of the force variation ratio for an entire sentence is a time-consuming, laborious job. Accordingly, a study was made of the consistency of the force variation ratio for adults and for a randomized sample of subjects in grades 4, 6, and 10 by correlating the force variation ratios for each word with the average force variation ratio for the entire sentence. These data are presented in Table III. The consistently high correlation between the force variation ratios for individual words and the average force variation ratio for the entire sentence may be noted. Since the force variation ratio for the words "fox" and "jumps" showed consistently the highest correlations, the sum of the force variation ratios for the phrase "fox jumps" only was used. In other words, the children wrote the complete sentence but an analysis of the pressure records of these two words only was made. As might be inferred from the data in Table III, the time required to write the phrase "fox jumps" correlated highly with the time required to write the entire sentence. Table IV demonstrates that the time for writing a short phrase is representative of the tempo of writing for the entire sentence.

Evaluations of the legibility of the handwriting samples produced by children and adolescents in this study were made by rating the legibility of the total sentence on an eleven point experimental scale composed of reproductions of the standard sentence, "The quick brown fox jumps over the lazy dog." This experimental scale was a composite scale made up of samples drawn from previously developed handwriting legibility scales prepared for grades 4, 6, and 10, respectively. The legibility ratings were done by five experienced judges who followed standard procedures in determining each child's legibility rating. Each rater displayed a relatively high degree of consistency in his own ratings. This is shown in Table V.

A generally high level of consistency among judges was also established. Median scale values for the ratings of the five judges were accepted as the best estimate of the legibility of the sample. In 122 cases out of 144 samples, the following criteria were met: 1) four or five of the ratings were identical with or contiguous to the median value (in 116 cases), and 2) a central tendency was clearly indicated (in six cases). The remaining 22 samples representing split distributions were then considered individually. For example, if one sample varied markedly from the majority of the ratings, it was re-rated. If the new rating did not move toward the median value, the atypical rating was discarded and the most typical retained.

Comparisons with the College Study—A summary of certain aspects of the methodology of the handwriting study of children and adolescents and that of the college study is given in Table VI.

It will be noted that the current study does not constitute a direct replication of the college study. In addition to obvious differences in the size and level of the population, other important specific differences should be noted:

1. The current study represents a randomized sample within purposely controlled factors of community, grade, and sex, while the college sample was stratified to identify the best and poorest writers in a given college population.

2. A single sample of handwriting was utilized in the current study, whereas subjects wrote a normal, a slow, and a fast sample in the college study.

3. In the current study the handwriting samples were rated on an 11-point scale, while in the

TABLE III

CORRELATIONS OF THE AVERAGE FORCE VARIATION RATIO FOR THE ENTIRE SENTENCE WITH THE FORCE VARIATION RATIOS OF INDIVIDUAL WORDS FOR 30 SUBJECTS RANDOMLY DRAWN FROM GRADES 4, 6, AND 10 AND FOR 19 ADULT COLLEGE SUBJECTS

| Word | 4th Grade (N=10) | 6th Grade (N=10) | 10th Grade (N=10) | Adults (N=19) |
|---|---|---|---|---|
| quick | .856 | .818 | .723 | .933 |
| brown | .809 | .824 | .847 | .919 |
| fox | .924 | .884 | .918 | .943 |
| jumps | .908 | .897 | .944 | .945 |
| over | .883 | .863 | .914 | .924 |
| the | .780 | .719 | .957 | .940 |
| lazy | .819 | .827 | .800 | .890 |
| dog | .914 | .804 | .927 | .956 |
| Σ "fox jumps" | .959 | .856 | .946 | .966 |

TABLE IV

CORRELATIONS BETWEEN WRITING TIME FOR THE PHRASE "FOX JUMPS" AND THE TOTAL SENTENCE WRITING TIME FOR 30 SUBJECTS RANDOMLY DRAWN FROM GRADES 4, 6, AND 10

| Grade | N | r |
|---|---|---|
| 4 | 10 | .962 |
| 6 | 10 | .963 |
| 10 | 10 | .847 |

### TABLE V

### CONSISTENCY CHECKS ON HANDWRITING LEGIBILITY RATINGS OF FOUR JUDGES* BASED ON 12 SAMPLES RANDOMLY DRAWN FROM A TOTAL OF 144SAMPLES

| | Deviation in Scale Steps from Original Rating | | | |
|---|---|---|---|---|
| | None | 1 | 2 | 3 |
| Rater 1 | 6 | 4 | 2 | 0 |
| Rater 2 | 11 | 1 | 0 | 0 |
| Rater 3 | 8 | 1 | 2 | 1 |
| Rater 4 | 7 | 1 | 4 | 0 |

*The 5th judge was unavailable when the consistency check was made.

### TABLE VI

### A COMPARISON OF THE METHODOLOGY USED IN THE HANDWRITING STUDY OF CHILDREN AND ADOLESCENTS AND THAT OF THE COLLEGE STUDY

| | Current Study of Children and Adolescents | Previous Study of College Students |
|---|---|---|
| Description of Sample | a. 144 subjects in grades 4, 6, and 10<br>b. 72 boys and 72 girls<br>c. A randomized sample of 12 boys and 12 girls at each of three grade levels in two communities | a. 19 subjects in junior year of college<br>b. 9 men and 10 women<br>c. A non-random sample of the best and poorest writers in a university class of 231 students |
| Number and type of handwriting samples | One sample (normal) per subject | Three samples (normal, slow, and fast) per subject |
| Method of rating legibility | Rating of normal samples on an 11-point composite scale. The scale was made up of samples drawn from Grades 4, 6, and 10. | Single rank-ordering of 57 samples (three samples per individual) of normal, slow, and fast writing |

college study samples were placed in rank-order of legibility. In both cases a satisfactory degree of consistency was obtained.

Differences in the methodology of the two studies produced certain differences in the data and likewise placed certain limitations upon comparisons of the data in the two studies. It seems clear that the random sampling technique in the current study tended to maximize the chances of a more normal distribution of handwriting legibility than was obtained in the college study. The use of a single sample of handwriting produced at a normal rate in the current study probably resulted in less variance in legibility than would have occurred had the samples been written at different speeds as in the college study.

Since the studies likewise differed markedly in the method of securing ratings of legibility, one using a handwriting scale and the other a rank-ordering technique, the college samples were later re-evaluated for legibility upon a scale and the relationships between legibility and the force variation ratio re-examined. These findings are reported in the next section of this paper, as well as the detailed findings of the relationship between handwriting pressure and legibility of handwriting in children and adolescents.

## Results of the Study

While the major emphasis of the present investigation was upon the determination of the relationship between the two factors, variation in point pressure and handwriting legibility, data were also obtained on writing rate and point pressure for boys and for girls at each of the three grade levels. In interpreting the results of the present investigation, reference will also be made to the findings of the previous college study, and to a follow-up study on a small sample of elementary school children in which the rate of writing was systematically varied.

Initial Steps in Treating the Data—It will be recalled that the subjects for the present investigation were randomly drawn from grades 4, 6, and 10 in the schools of two small urban communities. Data on handwriting legibility and variability in point pressure constituted the primary sources of data on the 144 children in these two communities. An analysis of variance design was employed to test the hypotheses that no true grade, sex, or community differences existed for either legibility or variability in point pressure.

The legibility variable was selected as the initial item for analysis. In selecting a level of confidence which seemed appropriate to the data and to the measuring techniques employed, the investigators set the one percent level. The results of the analysis of variance are given in Table VII. It will be noted that the F ratios of 43.18 and 19.08 for sex and grade respectively are significant beyond the .01 level, whereas no community differences in legibility were obtained. Hence, the null hypotheses were rejected for sex and grade, and the legibility data for the two communities were combined to determine what grade and sex differences in legibility did exist.

Sex and Grade Differences in Handwriting Legibility—Grade and sex trends in legibility including those for the college sample are shown in Figure 3. It should be noted that the slope of the legibility curve for the girls shows substantial improvement in legibility from grade 4 through grade 10, whereas such is the case for the boys. In fact neither the 10th grade boys nor the sample of college males achieved the mean legibility standard of the 6th grade girls.

Means and standard deviations for handwriting legibility are given in Table VIII based on the 11-point scale. While the means of the girls are superior to those of the boys at each grade level, the standard deviations are sufficiently large to assure considerable overlapping of the distributions at grades 4 and 6. In order to determine whether the observed differences in means by grade and sex were true or chance differences, the Duncan Range Test[4] was applied to the data for grades 4 through 10.

The results of this test showed that there were no significant differences in the mean legibility ratings between boys of different grades. Nor was the average legibility rating of fourth grade girls significantly different from the mean legibility ratings of boys in grades 4, 6, and 10. However, the test disclosed that the mean legibility of the sixth grade girls and the tenth grade girls differed significantly ($p = .01$) from each other and from each and all other mean legibility ratings for both sexes. In other words, the handwriting legibility of the tenth grade girls was superior to that of the sixth grade girls, and both the sixth and tenth grade girls were superior in legibility to the boys and girls at all other grade levels.

Difference in Force Variation Ratios for Sex, Grade, and Community—The data on variability in point pressure were analyzed in the same way as those for legibility. An analysis of variance was used to test the hypothesis that there were no true grade, sex, or community differences in variability in the application of point pressure. Table IX gives a summary of this analysis. As will be noted, a difference significant at the one percent level was found for grade only. However, sex and

---

[4] David B. Duncan. "Multiple Range and Multiple F Tests," Biometrics, XI (March 1955), pp. 1-9.

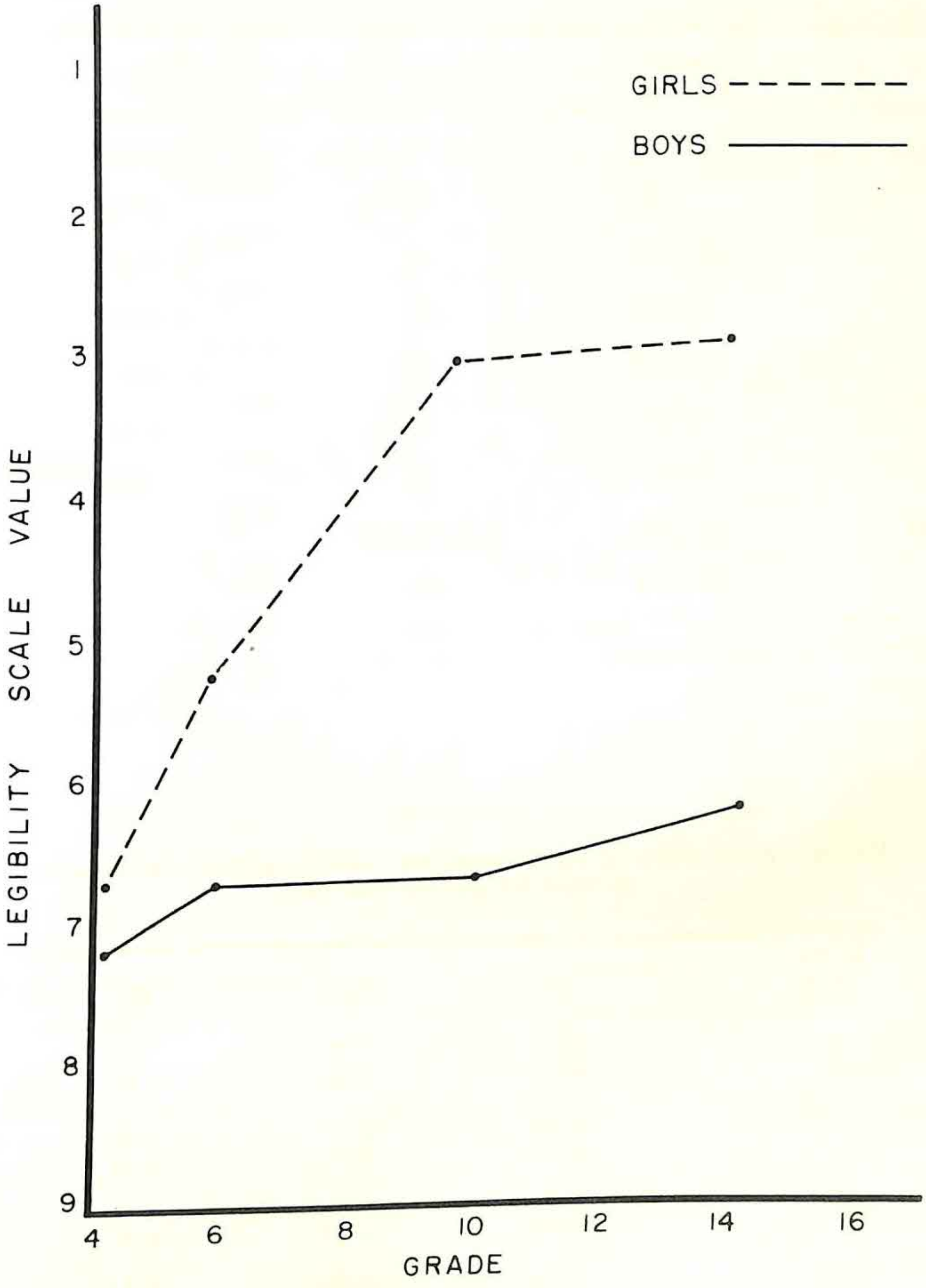


FIGURE 3: MEAN LEGIBILITY RATING BY SEX AND GRADE

### TABLE VII

### SUMMARY OF ANALYSIS OF VARIANCE FOR HANDWRITING LEGIBILITY

| Source of Variation | Sum of Squares | df | Mean Square | F |
|---|---|---|---|---|
| Sex | 115.5624 | 1 | 115.5624 | 43.1831* |
| Community | .1736 | 1 | .1736 | |
| Grade | 102.0972 | 2 | 51.0486 | 19.0757* |
| S × C | 1.5625 | 1 | 1.5625 | |
| S × G | 63.2914 | 2 | 31.6457 | 11.8253* |
| C × G | 3.3472 | 2 | 1.6736 | |
| S × C × G | 5.5418 | 2 | 2.7709 | 1.0354 |
| Within Groups | 353.2503 | (12)(11)=132 | 2.6761 | |
| Total | 644.8264 | 143 | | |

* Significant beyond the 1 percent level

### TABLE VIII

### MEANS AND STANDARD DEVIATIONS OF HANDWRITING LEGIBILITY SCORES BY GRADE AND SEX

| Grade Level | Sex | Mean | SD |
|---|---|---|---|
| 4 | Boys | 7.17 | .90 |
| | Girls | 6.83 | 1.73 |
| 6 | Boys | 6.71 | 1.46 |
| | Girls | 5.21 | 2.10 |
| 10 | Boys | 6.71 | 1.17 |
| | Girls | 3.17 | 1.86 |
| 12 | Men | 6.22 | 2.44 |
| | Women | 3.10 | 1.92 |

TABLE IX

SUMMARY OF ANALYSIS OF VARIANCE FOR FORCE VARIATION RATIO DATA

| Source of Variation | Sum of Squares | df | Mean Square | F |
|---|---|---|---|---|
| Sex | 23.2900 | 1 | 23.2900 | 4.6939* |
| Community | 23.3900 | 1 | 23.3900 | 4.7140* |
| Grade | 147.0966 | 2 | 73.5483 | 14.8229** |
| S × C | 14.8225 | 1 | 14.8225 | 2.9873 |
| S × G | 2.6689 | 2 | 1.3345 | |
| C × G | 8.6123 | 2 | 4.3062 | .8679 |
| S × C × G | 3.9631 | 2 | 1.9816 | |
| Within Groups | 655.9510 | (12)(11)=132 | 4.9618 | |
| Total | 878.7944 | 143 | | |

* Significant at 5 percent level
**Significant at 1 percent level

TABLE X

MEANS OF FORCE VARIATION RATIO BY GRADE AND SEX

| Grade Level | Sex | FVR Mean | FVR SD |
|---|---|---|---|
| 4 | Boys | 6.32 | 1.62 |
| | Girls | 6.84 | 2.20 |
| 6 | Boys | 6.49 | 1.52 |
| | Girls | 7.44 | 2.20 |
| 10 | Boys | 8.47 | 2.21 |
| | Girls | 9.23 | 2.85 |
| 15 | Men | 12.56 | 4.07 |
| | Women | 7.27 | 2.93 |

community differences were significant at the five percent level. So that the analysis of the data might parallel that for the legibility data, the results for the two communities were combined in further examining differences in sex and grade.

The mean force variation ratios by grade and sex are presented in Figure 4. It will be noted that with advancing grade level, the force variation ratio increases for both sexes through grade 10. The upward trend continues for college women.

Group data for the force variation ratios are given in Table X. Although the college scores were not included in the analysis of variance, mean scores for these subjects are also presented. The Duncan Range Test was applied to the force variation ratio data for grades 4 through 10. Significant differences were found between tenth grade girls and fourth and sixth grade boys and girls (p = .01), and between tenth grade boys and fourth and sixth grade boys (p = .01) and fourth grade girls (p = .05). While these grade differences are noteworthy, no true sex differences were obtained.

If Figures 3 and 4 are compared, it may be seen that the means in force variation ratio with advancing grade level are paralleled by an increase in handwriting legibility. Hence the relationship noted in the college study that highly legible handwriting is associated with a low force variation ratio is not supported by the data obtained on these children. In other words, with advancing age and improved handwriting legibility, one would have expected a decline rather than an increase in the force variation ratio. The data show, furthermore, that girls have a slightly higher mean force variation ratio at each succeeding grade level through grade ten than the boys while producing the more legible handwriting. At the college level, however, this trend is reversed, the women producing highly legible handwriting with low variability in point pressure. The men, on the other hand, continue the upward trend in variability in point pressure with a negligible increase in legibility. The reason for the contradictory findings of the two studies will become evident in a later section of this report.

Grade and Sex Trends in Writing Time—It is a common observation that young children write more slowly than older children, although there is considerable variability in writing rates among individuals of two given ages. The mean time for writing the standard sentence is plotted by grade and sex in Figure 5. It will be noted that the sex differences in rate of writing were not great although the mean writing time was less at successive grade levels. In fact the rate of writing more than doubled between grades 4 and 10. The most rapid gains in writing speed occurred between grades 4 and 6, with slightly less exaggerated gains between grades 6 and 10 and only a relatively slight increase in writing speed between grades 10 and 15.

Mean values for writing time by grade and sex are shown in Table XI. These data show that the mean writing time declined at a relatively constant rate with advancing grade level for both sexes. In general, the standard deviations indicate wide variability in writing times except for males in grade 10 and in the college sample. The Duncan Range Test was applied to the data for mean writing time for grades 4, 6 and 10. No significant sex differences within each of grades 4, 6 and 10 were found for mean writing time. However, the performance of fourth grade boys was significantly different that that of sixth grade boys (p = .05), sixth grade girls (p = .01), and tenth grade boys and girls (p = .01). The performance of fourth grade girls was also significantly different from that of sixth grade girls (p = .05), and from that of tenth grade boys and girls (p = .01). The mean writing time of the sixth grade boys was also significantly different from that of tenth grade boys (p = .05). These data clearly indicate that a substantial increase in speed of handwriting may be expected at successive grade levels, and that speed in handwriting appears to be a function of age rather than of sex.

Point Pressure Changes with Advancing Grade Level—During the initial states of writing there is a tendency for children to exert considerable point pressure. However, as children gain experience in writing general observation would lead one to believe that point pressure declines with age. To examine this observation, point pressure data were obtained from the oscillographic pressure records and are presented in Figure 6. The average force per second exerted by the point on the platen for the words "fox jumps" constituted the unit of measurement. It will be noted that with advancing grade level, average word force tended to decline through grade 10. However, sex differences in point pressure in grades 4 through 10 were slight. The college data for point pressure are not directly comparable and are therefore omitted.

Table XII shows that the variability among subjects in point pressure remained high at grades 4 and 6 even though the average writing pressure tended to decline. Only at the tenth grade were the standard deviations for writing pressure sufficiently low to indicate some uniformity among subjects in the application of point pressure. When the Duncan Range Test was applied to the data for point pressure, no significant sex differences were found within grades in mean point pressure. Significant difference (p = .05) were found for both sexes between grades 4 and 10 only. Thus it would appear that succesive grade changes for both sexes in mean point pressure are not as great as those in speed of writing.

Correlation Analysis—In order to determine the relationship between the force variation ratio

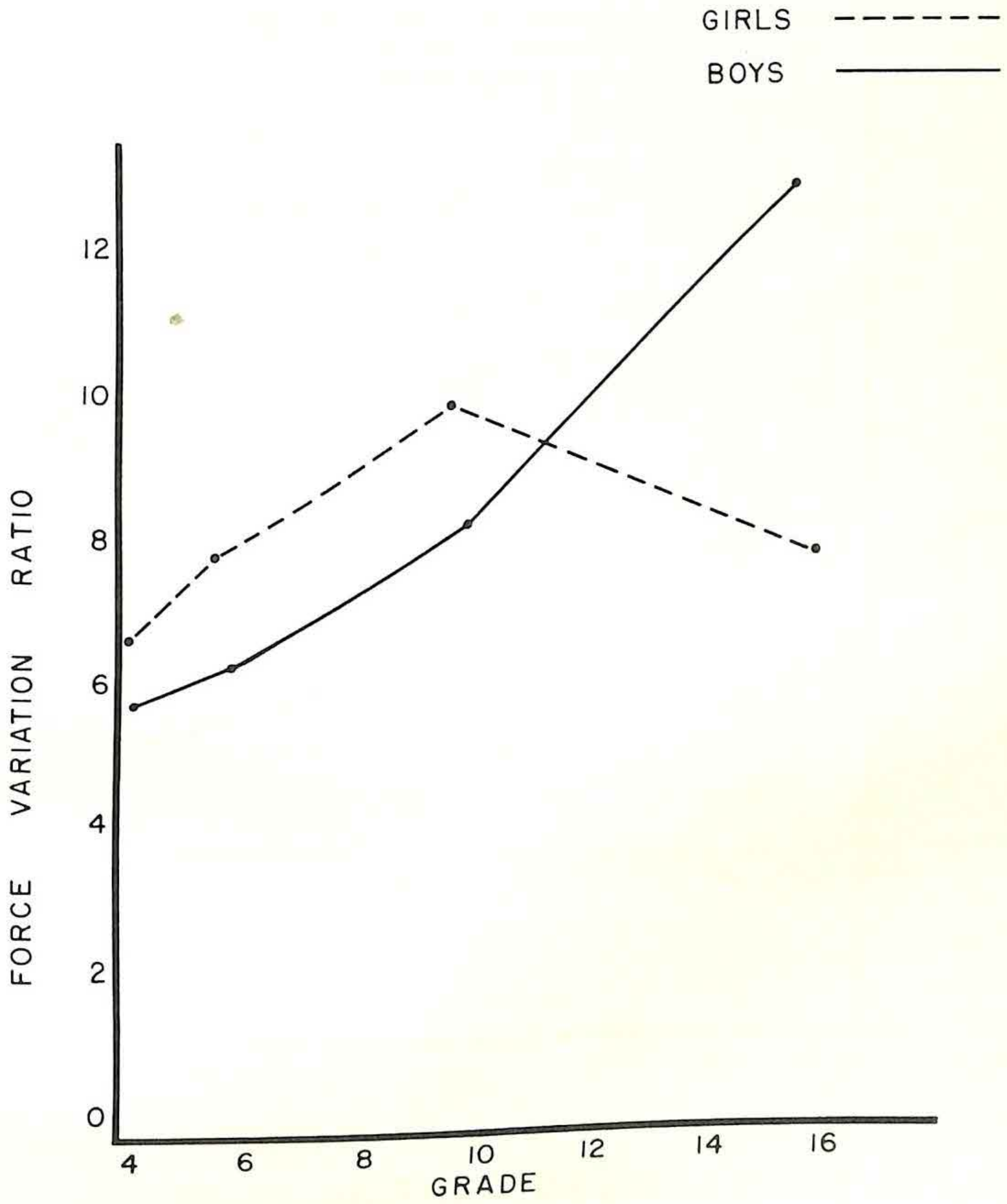


FIGURE 4: MEAN FORCE VARIATION RATIO BY SEX AND GRADE

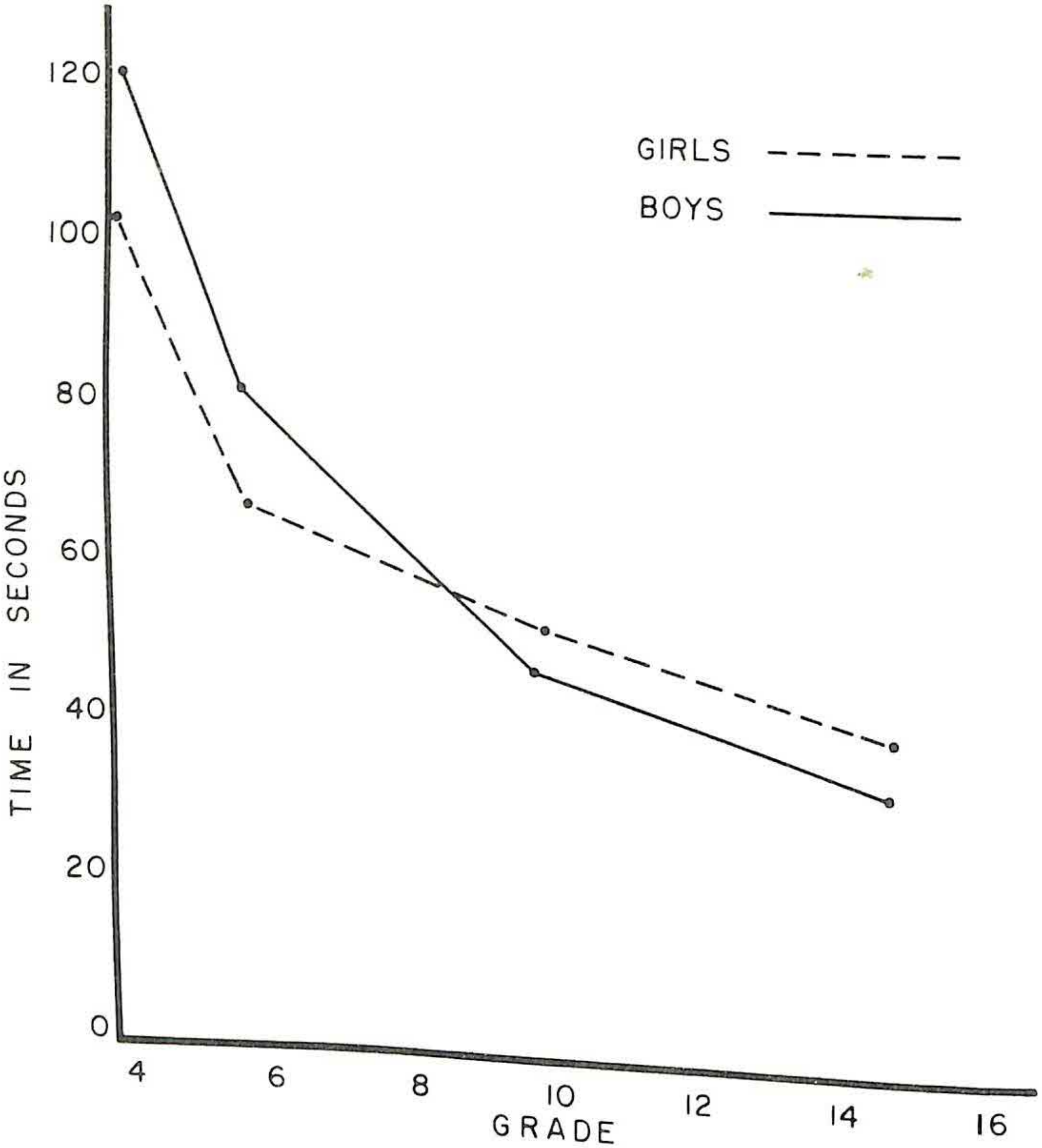


FIGURE 5: MEAN WRITING TIME BY SEX AND GRADE

TABLE XI

AVERAGE TIME REQUIRED TO PRODUCE THE STANDARD SENTENCE BY GRADE AND SEX

| Grade Level | Sex | N | Mean Time (in seconds) | SD |
|---|---|---|---|---|
| 4 | Boys | 24 | 119.53 | 39.91 |
| | Girls | 24 | 102.62 | 40.06 |
| 6 | Boys | 24 | 79.92 | 28.97 |
| | Girls | 24 | 65.64 | 14.74 |
| 10 | Boys | 24 | 42.02 | 6.81 |
| | Girls | 24 | 46.85 | 10.03 |
| College | Men | 9 | 32.12 | 2.18 |
| | Women | 10 | 39.47 | 8.09 |

TABLE XII

MEANS FOR POINT PRESSURE DATA BY GRADE AND SEX

| Grade Level | Sex | N | Mean Pressure (Av. height in mm.) | SD |
|---|---|---|---|---|
| 4 | Boys | 24 | 43.72 | 26.16 |
| | Girls | 24 | 41.69 | 20.96 |
| 6 | Boys | 24 | 30.80 | 10.61 |
| | Girls | 24 | 29.89 | 20.25 |
| 10 | Boys | 24 | 24.46 | 7.94 |
| | Girls | 24 | 24.88 | 7.29 |

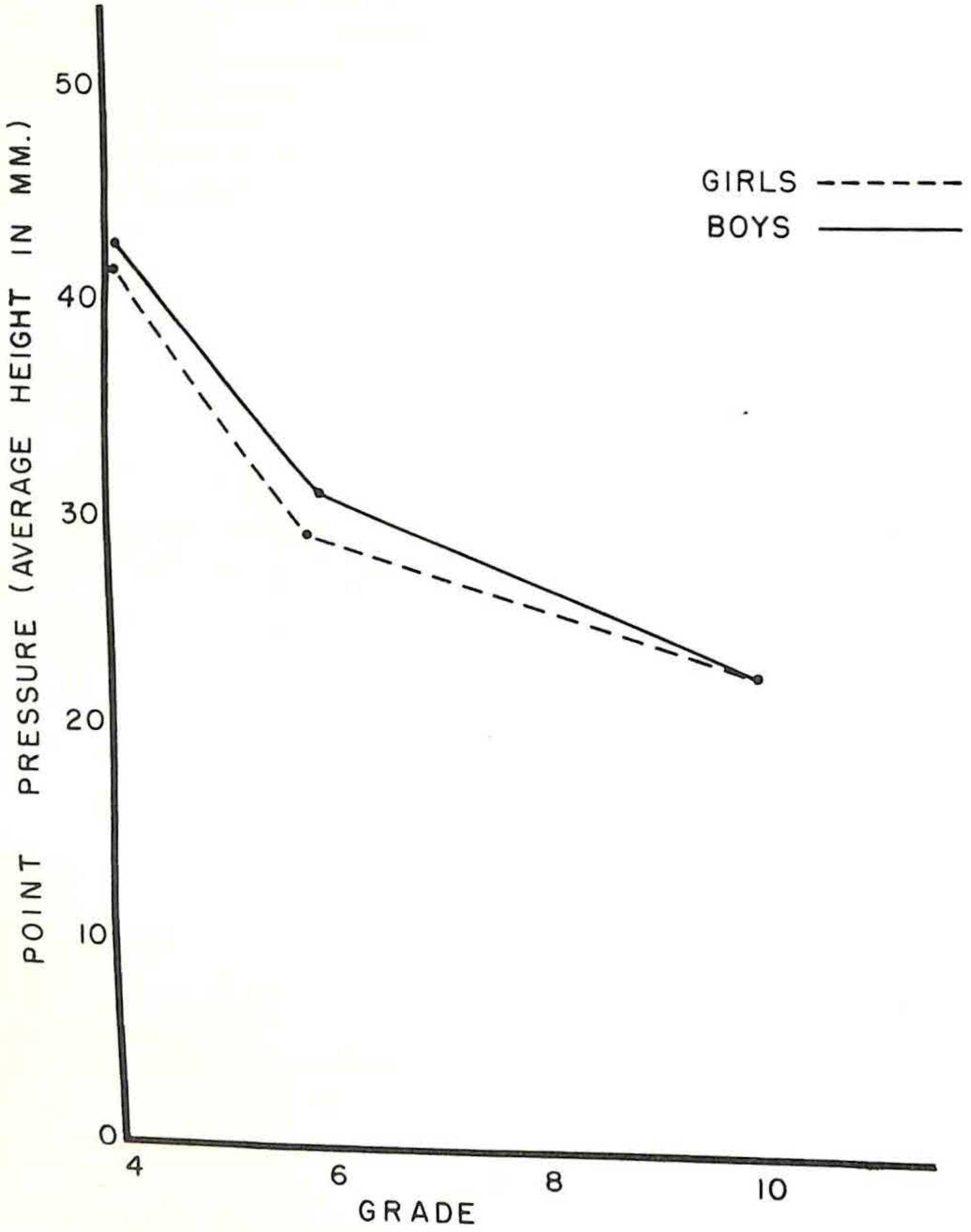


FIGURE 6: MEAN POINT PRESSURE BY SEX AND GRADE

and legibility, Pearson product-moment correlations were run between these two variables at the 4th, 6th, and 10th grade levels. These correlations are given in Table XIII. In order to interpret the apparent discrepancy between the data for the college population and the data for the present study, the correlations between the force variation ratio and legibility of the college population are also included.

It will be noted that all correlations between the force variation ratio and legibility ratings of the scaled samples for the 4th, 6th, and 10th grade pupils are low. It is evident, therefore, that under the conditions of this investigation, legibility and the force variation ratio appear to be unrelated. On the other hand, all correlations for the college subjects are positive and at least four are moderately high. Thus, when all 57 samples (slow, normal, and fast) for the 19 college subjects were correlated with the force variation ratio, moderately high correlations were obtained. The inclusion of the three samples at different speeds for each subject tended to broaden the range of legibility as well as that of the force variation ratio and hence tended to maximize the degree of correlation. On the other hand, when the analysis was restricted to only 19 normal samples, the effect of writing speed was minimized. Hence, the range of scores for legibility and for the force variation ratio was markedly reduced. This may, in part, explain the low correlations between these variables. It should also be noted that in those instances in which the samples were rated against the handwriting scales, lower coefficients of correlation were obtained than when the samples were ranked from most legible to least legible. In the present study, handwriting scales were used for all legibility rankings. It is believed that the use of scales tended to group the scores near the center of the distribution and yielded lower correlations than would have been found had the samples been placed in a single rank-order. This is illustrated by the low correlation of .18 found between the force variation ratio and legibility when the normal samples only for the college population were rated for legibility on the composite scale. This correlation for normal handwriting samples is comparable to that obtained for the children and adolescents. It is possible that had samples of normal, fast, and slow writing for children and adolescents been gathered and placed in rank order of legibility, correlations of a magnitude similar to those obtained for the mixed writing samples of the college population might have resulted.

Correlations Between Motor Control and Legibility—It will be recalled that upon completion of the standard sentence each pupil was given a battery of eight motor control tests (see Figure 2) which involved movements of the hand and fingers similar to but not identical to those used in handwriting. Pressure records were secured on all motor control items. Intercorrelations among the samples of force variation ratios for the eight items were consistently high, ranging from .609 to .948. The two items which correlated most highly with each other and with all others were the motor control items 40 and 10. As may be seen in Table XIV, the correlations between the force variation ratios (FVR's) for these two items ranged from .705 to .719. Since these two test items involve rather basic movements in letter formation and since the resulting FVR's are moderately highly correlated, it was decided to use the sum of the FVR's as the measure of motor control.

In order to examine further the stability of the force variation ratios, the variability in point pressure of the words "fox jumps" was correlated with the sum of the FVR's of the two motor control items. These correlations range from .353 to .539. It is interesting to note that while these correlations are not high, they are all positive and two of the three are significant at the 1 percent level and the other is significant at the 5 percent level.

The consistently positive correlations among these data suggest that these may be indicative of a definite motor function common to both tasks. The fact that these combinations are not as high as the correlation of .81 attained in a similar set of tasks in the college study may mean that a stabilized motor pattern for handwriting is not achieved until college age.

The correlations between legibility and motor control shown in Table XIV are negligible. These findings are in agreement with the correlations reported in Table XIII between legibility and the force variation ratio.

Legibility and Force Variation Ratio as a Function of Writing Speed—In order to study the effects which variations in the individual's rate of writing might have upon the relationship between legibility and variability in point pressure, data were obtained on another sample of 20 fourth grade children. These children wrote the standard sample at their usual rate of writing, at a slower rate than normal, and a rate which was substantially faster than normal. The results obtained for these 20 children are shown in Table XV. In arriving at the figures included in Table XV, an index value of 1.00 was taken as the average legibility and average force variation values at the normal rate of writing. Note that the values listed under the headings Slow and Fast are percentages of the mean values for the normal writing rate. For example, an index value of less than 1.00 indicates legibility better than that produced at normal writing rates, and a value higher than 1.00 is indicative of legibility poorer than that achieved at normal writing rates. Conversely, a variability index of less than 1.00 indicates relatively low force variability, whereas an index greater than 1.00 is

TABLE XIII

CORRELATIONS BETWEEN FORCE VARIATION RATIO AND LEGIBILITY RATING IN THE TWO HANDWRITING STUDIES

| Grade Level | N | Type of Sample | Correlations for Ranked Samples | Correlations for Scaled Samples |
|---|---|---|---|---|
| 4 | 24B | Normal | | -.21 |
| | 24G | Normal | | .25 |
| 6 | 24B | Normal | | -.10 |
| | 24G | Normal | | .16 |
| 10 | 24B | Normal | | .29 |
| | 24G | Normal | | -.17 |
| College | 57 | Mixed | .59 | .52 |
| College | 19 | Normal | .35 | .18 |
| College | 19 | Slow | .42 | .38 |
| College | 19 | Fast | .59 | .61 |

## TABLE XIV

## INTERCORRELATIONS OF MOTOR CONTROL DATA AND CORRELATIONS OF LEGIBILITY RATINGS WITH MOTOR CONTROL

| | 4th Grade N = 48 | 6th Grade N = 46 | 10th Grade N = 47 |
|---|---|---|---|
| FVR of ↺0 with FVR of ↻0 | .719* | .705* | .707* |
| Σ ↺0 ↻0 with Fox Jumps | .409* | .539* | .353** |
| Σ ↺0 ↻0 with Legibility Rating | .109 | .271 | .009 |

* Significant at the 1 percent level
**Significant at the 5 percent level

## TABLE XV

## LEGIBILITY AND FORCE VARIATION RATIO AS FUNCTIONS OF WRITING SPEED AT THE FOURTH GRADE LEVEL, USING THE AVERAGE FORCE VARIATION RATIOS OF THE NORMAL SAMPLES AS THE REFERENT

| | Slow | | | Normal | Fast | | |
|---|---|---|---|---|---|---|---|
| | Percent of Normal | | | | Percent of Normal | | |
| Legibility | Adult<br>Child | .77<br>.81 | high | 1.00 | Adult<br>Child | 1.55<br>1.29 | low |
| Variability in Force | Adult<br>Child | .80<br>.94 | low | 1.00 | Adult<br>Child | 1.52<br>1.62 | high |

evidence of higher force variation ratio than at the normal writing rate. Data from the college population were handled in the same way, using the averages of the samples produced at normal writing rates as the basis for computing indices for legibility and variability at slow and fast writing rates. It is interesting to note that complete agreement was obtained between the child and adult samples for both legibility and variability when the subject moved from the normal rate of writing to either the fast or the slow rate.

It is evident that when the legibility and the force variation ratio scores were examined as a percentage of the scores obtained at normal writing rates, high legibility was obtained on the average at slow writing rates and low legibility occurred when the writing rates were rapid. Likewise, when the writing rate was slow, variability in force for both groups of subjects was low and legibility was high; when the speed of writing was fast, variability in force was high and legibility was low. This lends support to the earlier observation about the college population that rapid writing tends to produce poor legibility and high variability in the application of force, whereas slow handwriting is associated with high legibility and low variability in application of force. It would appear, therefore, that the force variation ratio as such does not discriminate between the poor handwriter and the highly legible one at normal writing rates, but the evidence suggests that as a person moves away from his normal rate of writing, legibility and variability in force are related.

## Conclusions

This study has tested the hypothesis that legibility and force variation ratio are not significantly related in a child and adolescent population. The evidence presented in this study supported this hypothesis under conditions in which the children were asked to write at their usual rate of writing. While sex and grade differences in legibility were evident, these differences could not be attributed to differences in variability in point pressure. However, the data showed that when either children or adults moved away from their normal writing tempo, high legibility tended to be associated with low variability in application of force and poor legibility was associated with high variability in point pressure. In this sense, the original hypothesis was not supported for either children or adults.

The findings suggest that each individual has his own pattern for speed of writing and for variability in the application of force. These might be considered to be rather basic components of a motor set in handwriting. If speed of handwriting is increased, variability in application of force is likewise increased, the motor set is disturbed and handwriting legibility is adversely affected.

The Committee in Handwriting is now broadening the scope of its investigation to include other motor and perceptual aspects of the handwriting process. It is felt that the present study has provided additional evidence of the importance of focusing attention upon the individual and the stability of his performance under varying conditions. This would likewise appear to be a worthwhile focus for studying learning problems in other basic skills.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 1, September 1959)

# THE RUSSELL SAGE SOCIAL RELATIONS TEST: A TECHNIQUE FOR MEASURING GROUP PROBLEM SOLVING SKILLS IN ELEMENTARY SCHOOL CHILDREN*

DORA E. DAMRIN**
Educational Testing Service
Princeton, New Jersey

## Introduction

OVER THE past decade the development of new techniques of measurement in the general area of group dynamics has lagged behind the ever-increasing emphasis on research in this field. Although a few investigators have devised special methods for measuring some of the important elements of group interaction (1, 5), most researchers have, like Lippett and White, invented their own unique "test situations" and then employed the more familiar method of verbatim recording followed by detailed analysis (4). This paper describes an attempt which has been made to devise a "standard situation" in which defined elements of group interaction can be systematically observed, recorded, and scored.

The test is an outgrowth of the work done by the Mid-Century Committee on Outcomes in Elementary Education (2), and as such is geared to the measurement of social relations skills among elementary school children. It is believed, however, that the technique possesses wide applicability in areas other than the one for which it was originally devised, since it has been used successfully with teen-age as well as with adult groups in a variety of special fields. But, as described in the following pages, the test is presented as a possible research tool which might provide data to answer, at least in part, many of the important methodological questions which present-day educational techniques and practices raise.

## The Test and its Administration

The Russell Sage Social Relations Test is a situational test created to assess the nature and quality of two important aspects of elementary school children's skill in social relations: (1) skill in cooperative group planning procedures, and (2) skill in techniques of cooperative group action. It is designed for children in grades three through six and occupies approximately one hour of time. Its administration requires a trained examiner and a trained observer.

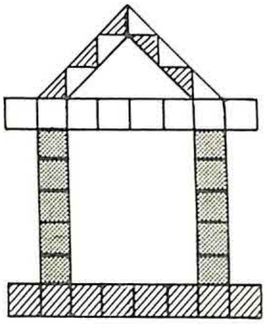
The test consists of three construction-type problems, graded in difficulty and administered one after the other. For each problem the children are provided with thirty-six interlocking construction blocks of various shapes and colors and a model which they are to copy exactly. All thirty-six blocks are necessary for the construction of each model, there being none left over when the problem is completed. As shown in Figure 1, the first problem requires the children to build the figure of a house; the second, a simple footbridge; and the third, the figure of a dog.

The test is administered to an entire classroom group of children at once. In the course of the administration, the examiner tells the children that the purpose of the test is to see how well they can work together. He tells them that the test does not provide scores for individual pupils, but only one score for the class as a whole. Pupils may, therefore, help each other, discuss things freely, and work cooperatively on the problem. He provides each child in the room with one or two of the blocks needed to build the particular model shown them. He then tells the group that, although they can have all the time they need to plan how they are going to go about it, they will be allowed only fifteen minutes for the actual construction. Finally, the examiner informs the group that their score on the test depends upon the time it takes

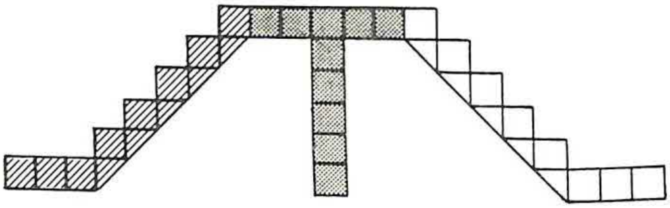
---

* The development of this instrument was made possible by a grant from the Russell Sage Foundation.

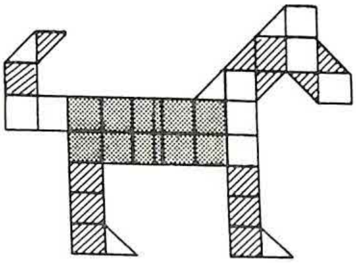
**The writer is indebted to Dr. William E. Coffman and Dr. William A. Jenkins for their fundamental contributions to the conceptual and theoretical framework of the test, as well as to its overall design.



Problem 1 — The house



Problem 2 — The footbridge



red ░

blue ▨

white □

Problem 3 — The dog

FIGURE 1

THE TEST PROBLEMS

them to build the model, the shorter the time the higher their score.

Having explained the test and its rules to the class, the examiner allows the pupils to proceed with the task of developing a plan of action. When this plan is completed and all members of the group have it clearly in mind, he tells them to begin construction of the first model and immediately starts timing. He stops his watch when the final block is correctly placed and announces the score to the class. He then proceeds with the second and more difficult problem. The class is again given time to plan and organize itself as in the first problem. The third and most difficult problem is administered in like manner.

As mentioned above, the test is designed to provide information about two different aspects of social relations skill. The first is the ability to participate in group discussions of plans for future group action; the second is the ability to regulate one's behavior in accordance with these plans so that movement toward the agreed-upon group goal is facilitated. The test is thus divided into two separate parts. The first part is termed the Planning Stage—the period during which the children devise a plan of action. The second part is called the Operations Stage—the period during which the group puts its plan into effect and actually builds the model, using the blocks provided by the examiner.

In both stages the aim is to obtain a measure of the best possible performance the group is capable of rendering. This aim is accomplished by having the examiner play a precisely defined and standardized role in his administration. His initial task is to explain the test and its rules to the children. In fulfilling this task his actions and statements are the same for all groups. His second and more complicated assignment is to create and maintain an atmosphere in which optimum group planning is possible. To achieve this purpose he must vary his behavior in accordance with the needs of the group. That is, a noisy group must be held in check so that communication and discussion of ideas is possible; a quiet solemn group must be encouraged to express ideas; a group that gets lost in the complexity of its own ideas must be rescued and its thinking clarified. Throughout all this the examiner must refrain from providing the children with his own ideas of ways of solving the problem and from telling them which of the ideas they have suggested are good or poor. He must at all times maintain a permissive atmosphere in which the giving of ideas is encouraged and all suggestions warmly accepted, but in which no positive direction is supplied. In such an atmosphere the children can perform only at the level at which they themselves are presently capable, and the quality of their planning thus provides a measure of the amount of skill which they possess.

The examiner's final task is to insure that, before entering the Operations Stage, each child in the group clearly understands the plan which has been devised. Therefore, he repeats verbatim the details of the plan which the class has built and asks if there are further questions before they begin the construction task. Here again the examiner must vary his behavior with the needs of the group, answering questions on points which the group has already decided, and referring back to the group any questions about points which they have not taken into account. It is not until the children tell the examiner that they are all ready to begin work that he permits them to start. This insures that all groups enter the Operations Stage with presumably equal clarity concerning the plan that they as a group have built.

During the Operations Stage the role of the examiner is also standardized. After the signal to start is given he withdraws to one side of the classroom and refuses to answer questions or provide help, even when requested. Regardless of what the children do (however noisy, rough, boisterous, or quarrelsome they become, or however quietly and efficiently they work), he neither reprimands nor commends them. He interferes only if behavior gets so out-of-hand that physical harm to children is apt to result. If this occurs, he stops the test and no further problems are given.

The mechanics of administering the test place the children in a situation where progress toward the goal is not possible without some type of cooperation on the part of each child in the group and where external controls designed to elicit particular kinds of behaviors and suppress other kinds do not exist. In such a situation the children can show the extent to which they are capable of carrying out a plan of action they themselves have devised, and the kind of behavior they exhibit provides a measure of the amount of skill they possess in this respect.

During the time the children are engaged in the task of planning how to solve the problems and in working on the actual construction of the various models, the observer (seated near the back of the classroom) keeps a record of their behavior using standardized observation sheets. The data thus collected are later transformed into a set of numerical scores which are used to rank groups on the basis of their skill in cooperative group planning procedures and in techniques of cooperative group action. (A manual containing complete directions for administering and scoring the test can be obtained from ETS.)

## Empirical Development of Scoring Procedures

The methods of scoring which are discussed in a subsequent section are the result of two years' research. Table I shows the large-scale administrations that have been completed, but fails to il-

TABLE I

EXPERIMENTAL ADMINISTRATIONS OF THE RUSSELL SAGE SOCIAL RELATIONS TEST

| Date | No. of Participating Schools | Grade | No. of Classes |
|---|---|---|---|
| November 1953 | One | Third | 4 |
| | | Fourth | 4 |
| | | Fifth | 4 |
| February 1954 | One (retest) | Third | 4 |
| | | Fourth | 4 |
| | | Fifth | 4 |
| April 1954 | Fifteen | Fourth | 10 |
| | | Fifth | 10 |
| | | Sixth | 10 |
| November 1954 | Twenty-one | Third | 24 |
| | | Fourth | 13 |
| | | Fifth | 9 |
| | | Sixth | 3 |
| May 1955 | Twenty-one (retest) | Third | 24 |
| | | Fourth | 13 |
| | | Fifth | 9 |
| | | Sixth | 3 |

lustrate the fact that after each administration the observation, recording, and scoring techniques employed underwent drastic revisions. Each revision represented yet another attempt to break away from the familiar framework of scoring individuals and to devise a reliable and meaningful system for scoring groups.

Such attempts proved to be unusually difficult since all previous research in this general area has employed techniques which describe segments of individual behavior, either by means of a check list such as that devised by Bales (1) or by means of verbatim recordings of individuals' comments, attitudes, actions, etc., as employed by Lippitt and White (4). All attempts to apply such techniques proved unsatisfactory from the start. In the first place, the groups dealt with were so large that in focusing upon the behavior of specific individuals the interaction among individuals was lost sight of. Secondly, the varieties of interaction which occurred in different groups were so complex that individual recordings failed to catch much of the important information. Finally, a primary aim was to develop an instrument which teachers themselves could use with their classes; hence complex techniques which required an undue amount of time, effort, special training, and expense could not be used.

Further search of the literature revealed that most research on groups of the size in which we were interested used sociometric ratings and specially designed questionnaires. The few special techniques that had been devised were similar to Pepinsky's Group Participation Scale (5), a modified "guess-who" instrument which not only depended heavily upon the intelligent cooperation of the subjects but produced a limited picture of only one aspect of group interaction.

Two important guideposts determined the general direction which the development of observation, recording, and scoring techniques for the present test should take. The first of these was the definition of the "good" group implicit in the Social Relations section of Elementary School Objectives (2). The second was the general theory of behavior outlined by Krech and Crutchfield (3), particularly those parts of the theory which related to frustration and aggression. We discovered that "good" groups, although few in number, did exist and were readily distinguishable from "not good" groups. To extract the simple variables that reflected these differences in a precise manner proved a lengthy and time-consuming task.

The November 1953 administration was conducted to see whether the test would "work" with elementary school children. The results were quite satisfactory, children being both interested in and motivated by the problems. Following this administration a beginning was made on the development of recording techniques which would successfully capture, and scoring techniques which would accurately reflect, the wide differences in the test behavior of different groups. Tentative procedures were devised and tried out in a retest of the same groups in February 1954. Although this retest pointed up innumerable flaws in the methods so far developed, sufficient data were accumulated to show that over this three-month interval some of the groups had improved their test performance, whereas others had deteriorated noticeably. A search for the reasons which lay behind such changes led to the formulation of the following hypothesis: changes in test behavior over time are a function of: (1) the nature of the teaching methods a teacher employs in her classroom, and (2) the kinds of controls the teacher uses to make her methods effective.

To test this hypothesis, the cooperation of the elementary school supervisor in a large city system was solicited. The supervisor was made acquainted with the test and with our ideas regarding what it measured and was asked to select for us five pairs of teachers at each of three different grade levels, fourth, fifth, and sixth. One member of each pair was to be a teacher who placed great emphasis upon teacher-pupil planning and group work. The other member of the pair was to be a teacher who placed major emphasis upon individual instruction and who did not use group planning techniques with her children. Since we had no advance knowledge as to which member of a pair was which, we felt that if the social relations test could discriminate between the two it would indicate that our hypothesis was a tenable one. Of the thirty classes tested in the April 1954 administration, twenty-four were predicted correctly. The six incorrect predictions were found to be due in large part to errors and ambiguities which still existed in our recording and scoring procedures. Accordingly, all procedures were once again subjected to a few major and several minor revisions. The revised techniques were pretested on classes in several near-by county schools and were found satisfactory.

The final step was to determine if others could be trained to use the test and obtain results which corresponded to ours. To answer this question we accepted an invitation to participate in a comprehensive teacher evaluation study being conducted by the Division of Teacher Education in New York City. The fifty classes tested in the November 1954 administration and retested in May 1955 provided opportunity to experiment with the training of four different teams of examiners, one administrator and one observer per team.

The November administration was disheartening, for we found that several of the behaviors to be observed and recorded were still not defined with sufficient objectivity to render them reliably scorable by persons not intimately acquainted with the test. The retest in May, however, proved more successful, probably because between the

November and May administrations much effort was devoted to spelling out in more precise behavioral terms the variables we were interested in. It was found that a person with the equivalent of a bachelor's degree in education who has had some teaching experience can be trained to administer the test in approximately two days, working with two or three classes each day. Observers, however, require approximately four or five days of training at the rate of at least two classrooms per day. Funds were not available for having two observers in each classroom, hence reliability data could not be collected. The writer worked with each observer until an estimated reliability of .80 was obtained on at least two classes. The observer was then on his own.

Despite the fact that continuous experimental revisions prevented our obtaining and reporting a series of coefficients descriptive of the reliability and validity of the test, it is felt that valuable evidence of a nonstatistical and subjective type can be reported.

1. Variations in class size ranging between 20 and 30 seem to have no effect upon children's test behavior. If a group is larger or smaller, test behavior apparently is altered in direct proportion to the amount of departure in size from this range. For example, communication is much easier in a group of 6 or 7 persons than in a group of 40 or more persons, but of approximately equal difficulty in groups of 20 and groups of 30.

2. Persons with some educational background can be trained easily to administer the test using the standardized procedures which are necessary for adequate and reliable assessment of group behavior.

3. Definite evidence has been collected showing that children's behavior in the test situation reflects in some degree the nature and quality of the educational experiences provided by the teacher. This finding not only supports, but is directly in line with, the results of the Lippitt and White experiments on the social climate of children's groups (3).

4. Tentative evidence has been collected in support of the hypothesis that the administrative set-up of a school is a more important determinant of children's test behavior than the classroom methods employed by a particular teacher. The scores of classroom groups of children seem to vary less within schools than they do between schools. A school with a warm, friendly principal, whose teachers appear to have a great deal of individual freedom and good inter-staff relationships, tends to have classrooms which score relatively high on the test. A school in which principal and teachers are stiff, cold, fearful of the test itself, and aggressive toward the examiners tends to have classrooms which score relatively low.

5. Retesting children after a lapse of from three to six months and using the same problems has been found to be a feasible procedure. The second administration is always given with the instruction, "We would like to see how much better you can do this time," and with but rare exceptions children enter into the task as eagerly and as cooperatively as they did at first. Changes in behavior from one administration to the next apparently reflect the nature of the classroom learning experiences children have been exposed to in the interim, rather than learning how to take the test.

6. Quite by accident it has been found that in many ways the test is as applicable for adult groups as for children's groups and can, with appropriate modifications, be scored in the same way. We have found, for example, that an adult group unskilled in techniques of cooperative group planning scores definitely below a third-grade group which has received training along these lines, and that an adult group highly skilled in such techniques surpasses a highly skilled children's group only in the complexity and sophistication of their final plan of action.

The problem of scorer reliability, however, is still unsolved and it is felt that, before the techniques described in the following section can be used with confidence, they must be subjected to a detailed and rigorous statistical analysis. Research on this problem is now in progress.

## Recording and Scoring Procedures

Since the types of scores a situational test provides are determined by the specific behaviors which are observed and recorded, a first step in scoring is that of setting up behavioral definitions of the skills the test is designed to measure.

On the basis of the material presented in the section on Social Relations in Elementary School Objectives (2), as well as on that found in the literature of education and of group dynamics, a group skilled in cooperative planning techniques can be said to be one characterized by a widespread, interested exchange of appropriate ideas in which definite growth in the direction of greater clarity and precision of thinking occurs as the discussion progresses. In addition, the skilled group possesses a substantial amount of autonomy in that the members can, without external help or restraint, conduct themselves during the discussion in such a way that progress toward their goal of devising a mutually satisfactory plan of action is implemented rather than impeded. A group skilled in techniques of cooperative group action can be said to be characterized by a harmonious working atmospere in which there is widespread and interested concern on the part of all members regarding rapid and efficient progress toward the group goal.

Using these definitions as guides, two sets of relatively independent variables (one for the Planning Stage and one for the Operations Stage) were developed for scoring purposes. For the Planning Stage the variables are:

1. Participation—The extent to which individual children in the group enter actively into the planning discussion. In a skilled group many children participate; in an unskilled group few children participate.

2. Involvement—The extent to which children exhibit interest in and concern for the task at hand. In a skilled group the majority of the children evidence an active and sustained interest in the planning. In an unskilled group interest is quite low and may be completely lacking.

3. Communication—The way in which children exchange ideas during the discussion period. In a skilled group children listen to each other, critically and constructively evaluate each others ideas, and integrate simple ideas to build improved and more complex ones. In an unskilled group children act independently, and there is no real or fruitful exchange of ideas and suggestions.

4. Autonomy—The extent to which children can discuss the problem and reach a point of decision about a final plan without the help or restraint of the examiner. The skilled group is almost completely independent of the examiner. The unskilled group requires continuous help and/or restraint to get through the planning period.

5. Organizational Techniques—The kinds of ideas children have about ways of organizing themselves to attack the problem. In a skilled group the major part of the discussion is centered around concepts of leadership and organized subgroups. In an unskilled group the discussion centers around the general idea of having each child work as an individual.

6. Final Plan—The quality of the plan which the children eventually devise. In a skilled group the plan is precise and detailed, with every child in the group knowing exactly where he is to go and what he is to do. In an unskilled group the plan is essentially nonexistent, being so vague and indefinite that no child in the group has any clear idea of what his responsibility, as an individual, is with regard to building the model.

For the Operations Stage the variables are:

1. Involvement — The extent to which children exhibit interest in and concern for accomplishing the task at hand. In a skilled group the majority of the children evince sustained interest in the progress being made either by themselves or by those who are engaged in doing the building. In an unskilled group many children withdraw from the field altogether and pursue non-problem-centered activities.

2. Atmosphere—The psychological tone-quality of the group of children who remain in the problem field as revealed by the kinds of statements the children make and by the tone of voice in which the statements are made. In the skilled group, problem-solving is carried on in a warm, friendly, supportive, and harmonious atmosphere. In the unskilled group the atmosphere can range from an extreme and unnatural quiet to screaming excitement or open fighting.

3. Activity—The kind of behavior children who have withdrawn from the problem field engage in. In the skilled group few or no children withdraw, and the few who do engage in acceptable classroom behaviors. In the unskilled group many children tend to withdraw, and the majority of them participate in highly unacceptable classroom behaviors.

4. Success—The time necessary for the group to complete the task of building the model. The skilled group can complete the task in ten minutes or less. The unskilled group seldom finishes within the allotted fifteen minutes, or their behavior becomes so out-of-hand that the test must be stopped before the fifteen minutes are up.

Space does not permit a detailed report of the standardized directions which have been developed concerning the specific behaviors which are observed, the way in which they are recorded, and the way in which they are combined to provide numerical scores on each of these ten variables. For present purposes it must be sufficient to state that on the basis of the data recorded on the observation sheets every group taking the test receives a score of from 1 to 5 on each variable, 1 representing the poorest score and 5 representing the best score. These scores are then combined to yield a single descriptive rating which provides a meaningful picture of the group's social relations skill.

In the beginning phases of the research an attempt was made to combine the six variables of the Planning Stage and the four variables of the Operations Stage in such a way that groups could be ranked from high to low along each of two continua designated as skill in cooperative group planning and skill in cooperative group action. It was soon found, however, that a simple ranking of this type disregarded much of the information that was deemed educationally and psychologically significant. The reason was an obvious one. Skill in cooperative group planning does not consist of one but rather of several interrelated and interdependent component skills. The same is true of skill in cooperative group action. Groups can (and do) vary from each other in any one, two, three, or all of these component skills as well as in different combinations of them. It became essential, therefore, to devise a conceptual framework for describing differences in the behavior of groups that could handle variations in several variables rather than variations in a single variable.

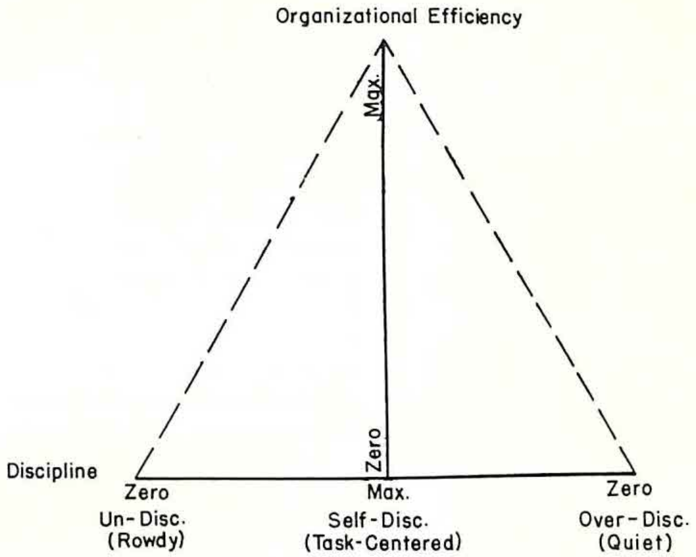
Study of the test performance of many different groups of children revealed that in each stage of the test certain variables acted as limiting factors on others. These factors, termed limiting variables, were of such a nature that if a group received low scores on them the group was prevented from receiving high scores on any of the other non-limiting variables. The converse was not true. Groups who received high scores on these limiting variables did not necessarily receive high scores on the non-limiting ones, but could range in score from very poor to very good. This finding eventually led to the postulation of two axes, one composed of the limiting variables and the other composed of the non-limiting ones. These axes were then aligned in such a way that a group's position on the one determined the range of positions it could occupy on the other.

Examination of the psychological nature of the variables which determine a group's position on these two axes revealed that the limiting variables were related to the amount and kind of discipline children exhibit in the test situation. The non-limiting variables concerned the amount of technical knowledge children seemed to possess about how to organize themselves to accomplish a group task. The reason for the somewhat peculiar relationship between the several variables which go to make up skill in social relations now becomes clear. The technical knowledge which members of a group possess cannot be put to use unless members also possess an adequate amount of self-discipline. That is, children (and adults too, for that matter) might be aware of what "ought" to be done in order best to accomplish a specific group goal, but, unless each individual conducts himself in a manner conducive to such accomplishment, optimum progress toward the goal becomes impossible. On the other hand, effective self-discipline on the part of group members cannot, in and of itself, serve to make up for lack of the technical "know-how" necessary for maximum group efficiency. The extent to which a group is able to utilize effectively the technical knowledges and skills which its members possess is thus limited by the extent to which the members are self-disciplined.

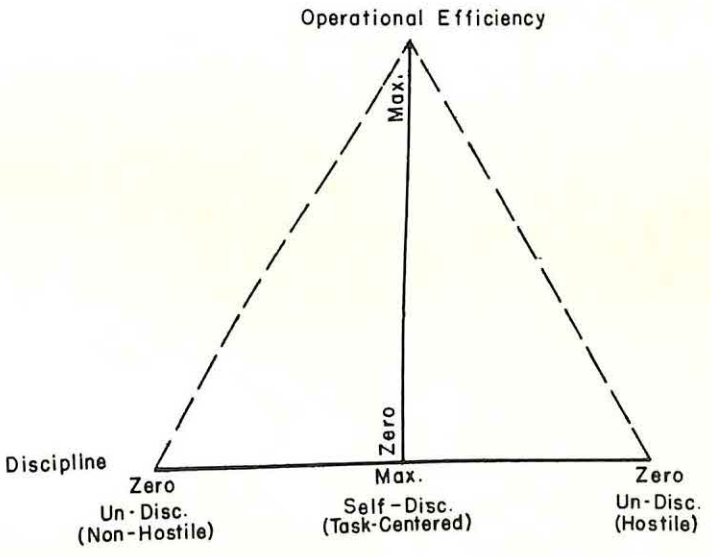
Figure 2 shows the way in which two axes are aligned in the Planning Stage and in the Operations Stage of the test. The horizontal axis has been designated Discipline, and a group's placement on it is determined by scores received on the limiting variables in each stage. The vertical axis which bisects it has been designated Efficiency (Organizational Efficiency in the Planning Stage and Operational Efficiency in the Operations Stage). A group's placement on this axis is determined by the non-limiting variables in each stage. The dashed restraining lines in the figure are purely illustrative, serving to indicate that a group's distance in either direction from the mid-point of the horizontal axis limits, but does not wholly determine, the height to which the group can rise on the vertical axis.

In the Planning Stage the variables Involvement and Autonomy were found to be limiting ones. Involvement, it will be recalled, is a measure of total group interest, and Autonomy is a measure of the extent to which a group can function independently of the examiner. Only a group whose members possess a high degree of self-discipline has the ability to get through the planning discussion without examiner interference. Group members will maintain interest in the planning discussion only to the extent that they feel themselves to be needed, wanted, and important members whose ideas are worthy of attention and consideration by others. The more this feeling is present, the more self-disciplined a group tends to be and the more it is able to function independently. Similarly, the more self-disciplined a group is, the more its members tend to consider, as well as seek, the ideas and opinions of others. Autonomy and Involvement are thus seen to be inextricably related, and without these two conditions the profitable exchange of ideas necessary to good group planning is impossible.

As Figure 2 suggests, a group's lack of Autonomy in the Planning Stage can manifest itself in different ways. Groups can be un-disciplined in the sense that all forms of control, both external



THE PLANNING STAGE



THE OPERATIONS STAGE

FIGURE 2

SCHEMATA USED TO SUMMARIZE THE DATA OBTAINED FOR GROUPS IN THE PLANNING AND OPERATIONS STAGES OF THE TEST and internal, are absent, or they can be over-disciplined in the sense that external controls have been so strongly imposed that the development and/or functioning of self-discipline is not possible. In the test situation the completely undisciplined group is a rowdy, noisy, restless one in which children pay little or no attention either to the examiner or to the task at hand. The extremely over-disciplined group, on the other hand, is one in which children are so restrained and passive that the examiner can elicit almost nothing from them. Both groups, in the extreme case, are almost totally lacking in interest and involvement in the test problems, and this combination of factors prevents them from functioning efficiently during the planning period.

In the Operations Stage of the test the limiting variables were found to be Atmosphere and Activity. Atmosphere refers to the psychological tone-quality of the group of children who remain in the problem field and who exhibit concern for the construction task. Activity refers to the kind of behavior children who have withdrawn from the problem field engage in. A harmonious working or problem-solving atmosphere in which optimum progress is possible can, in the present test situation, be created and maintained only if each group member subjects himself to the rules of the plan devised during the discussion period and refrains from engaging in any behavior which will impede progress toward the group goal. Group members who have withdrawn from the problem field must likewise refrain from engaging in behaviors which are so rowdy or so hostile that they disrupt the work in progress. Atmosphere and Activity, therefore, are concerned with the amount of self-discipline group members possess, and directly affect the efficiency with which the group as a whole can accomplish its task.

Lack of self-discipline in the Operations Stage results in one of two different types of behavior, both of which are equally detrimental to efficient group accomplishment. As shown in Figure 2, children who are undisciplined can evince hostile or non-hostile behaviors. In the extreme case the undisciplined non-hostile group is one in which the classroom is turned into a playground, children running about with complete abandon. On the other hand, the group in which members exhibit extreme hostility towards each other is one in which angry quarreling and open fighting predominate. In both groups operational efficiency is so low that it frequently is necessary to stop the test within the first few minutes.

All four of the non-limiting variables which determine a group's placement on the Organizational Efficiency line in the Planning Stage relate to those technical skills and knowledges essential to building a good plan of action. Since each child in the room has one or two of the blocks necessary to build the puzzle, it is obvious that some type of organization is essential to efficient accomplishment of the construction task. If children have no knowledge of organizational techniques such as those of delegating leadership, appointing subgroups to perform certain tasks, or specifying when and where different groups are to work, a good final plan cannot be built. Possession of such knowledge, however, is not by itself sufficient. Children must also possess the ability to express their ideas, to communicate them to others, to listen to the comments and criticisms which others might make, and to engage in that fruitful exchange of ideas and points of view which results in the most precise, the most comprehensive, and the most detailed plan of action. The extent to which group members possess these knowledges and skills is reflected by scores on the variables Communication, Participation, Ideas, and Final Plan.

In the Operations Stage the non-limiting variables are Success and Involvement. Success is a highly objective measure of efficiency based upon the time it takes the children to build the model; it is dependent upon both the excellence of the plan and the discipline with which it is carried to completion. Involvement, although still a measure of total group interest, is not a limiting variable in this stage as it is in the Planning Stage. This is because skill in techniques of cooperative action is revealed in the extent to which children, when placed in a completely free and unstructured situation, can exhibit and maintain genuine concern for accomplishing a group goal. This concern may be manifested directly or indirectly. Direct Involvement is exhibited by children who remain in the vicinity of the problem table, following closely the progress being made on the task, and also by children who, although drawing slightly apart from the builders, obviously keep an eye on the progress being made. Indirect Involvement is evidenced by children who, quietly confident that in the Planning Stage they, as a group, have made the best possible plan of action, pursue other ongoing classroom activities and leave those few who have been delegated to perform the construction task to carry out their responsibility.

It is possible for a group to receive a high score on Success (indicating that the problem was solved in a relatively short period of time) because the majority of the children refused to accept the problem and withdrew from the field altogether, leaving the job in the hands of three or four group members. In such cases the problem is obviously no longer a "group" problem and the total group's rating on Operational Efficiency is low despite their high Success score. Success, measured in number of minutes it takes to solve the problem, thus becomes a meaningful measure of "skill in cooperative group action" only when the Involvement score is high, indicating that the majority of group members remained interested in the accomplishment of the task. For this reason, Operational Efficiency must be determined by scores on both Success and Involvement variables.

To summarize, a group's placement on the discipline axis is determined in the Planning Stage by scores on Autonomy and Involvement, and in the Operations Stage by scores on Atmosphere and Activity. Placement on the Organizational Efficiency continuum is determined by scores on Participation, Communication, Ideas, and Plan, and on the Operational Efficiency continuum by scores on Involvement and Success.

### Classification of Groups on the Basis of Test Behavior

Every group that takes the test is classified on two bases: (1) their Planning Stage behavior, and (2) their Operations Stage behavior. This classification is made using the theoretical framework described above for summarizing the data obtained for a group. For ease of communication, each "group-type" has been assigned a descriptive name, as shown in Figures 3 and 4. The dividing lines in these figures are by no means rigid, serving only to illustrate the fact that the defined areas represent points of concentration in a plane. All of these group-types have been observed and recorded empirically. Their major characteristics are given below.

### Group-Types in the Planning Stage

Figure 3 shows the seven major types of groups that have been defined for the Planning Stage of the test. The characteristics of each type is as follows:

1. The Mature Group—Characterized by good ideas about ways of organizing themselves and by an excellent communication pattern. The discussion is widespread, animated, self-controlled and well conducted. Frequently the examiner can retire altogether, leaving the total planning session in the hands of one or two student leaders.

2. The Dependent Group—Characterized by good ideas about ways of organizing themselves and by a good communication pattern. This group, however, possesses less autonomy than the Mature Group, and to get them through the discussion the examiner has to provide frequent assistance in the form of summarizing ideas, suggesting next steps, reminding them to speak one at a time, etc.

3. The Immature Group—Characterized by an alert and eager interest in the test, combined with an almost total lack of skill in group problem-solving procedures. Ideas about organizational techniques are to all practical purposes non-existent and the group generally accepts by consensus the first or second idea mentioned, regardless of its feasibility. The active interest, however, serves to keep the group well disciplined and little in need of examiner aid or restraint.

4. The Semi-Controlled Group—Characterized by fairly wide and generally interested participation. The communication pattern is relatively good, as is knowledge of organizational techniques and quality of the final plan. Group discussion is markedly impeded by the fact that group members tend to talk all at once. The examiner is always hard put to keep the classroom sufficiently quiet and the discussion sufficiently centralized so that children can hear each other and really develop their ideas.

5. The Semi-Restrained Group—Characterized by an atmosphere of "dutiful recitation." Children are quiet, polite, attentive, and respond readily to the examiner's requests. Their ideas are generally good, but the planning discussion tends to be stiff and artificial, as though children were doing what they feel they "ought" to do rather than entering freely, wholeheartedly, and with enjoyment into the task at hand. The discussion is seldom widespread; it is generally carried by one, two, or three children, the rest remaining quietly passive and disinterested.

6. The Uncontrolled Group—Characterized by a complete lack of discipline, combined with a total lack of skill in techniques of planning and communication.

7. The Restrained Group—Characterized by an almost complete and highly unnatural silence, combined with zero skill in techniques of planning and communication. The examiner must work very hard with this group, prodding them constantly to get them to respond at all.

### Group-Types in the Operations Stage

In the Operations Stage, as in the Planning Stage, the results of many administrations of the test have shown that groups tend to cluster in a few areas along the two axes, Discipline and Operational Efficiency. As shown in Figure 2, lack of self-discipline in this stage of the test can be evidenced in either hostile or non-hostile behavior. The type of behavior exhibited serves to locate a group on the right or left side of the tri-

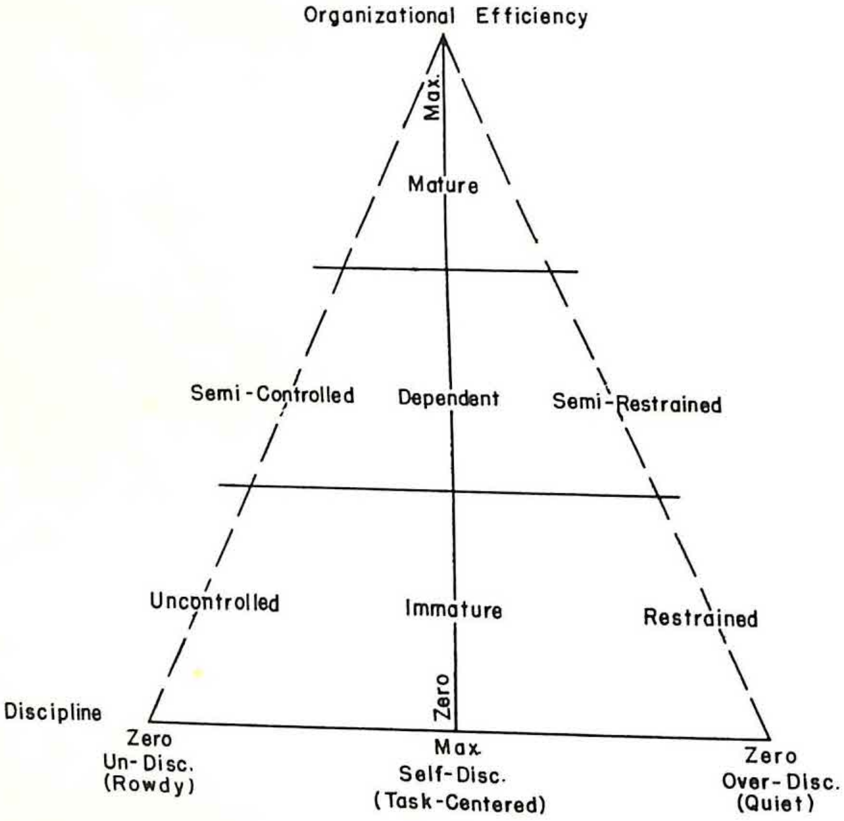


FIGURE 3

TYPES OF GROUPS WHICH OCCUR DURING THE PLANNING STAGE

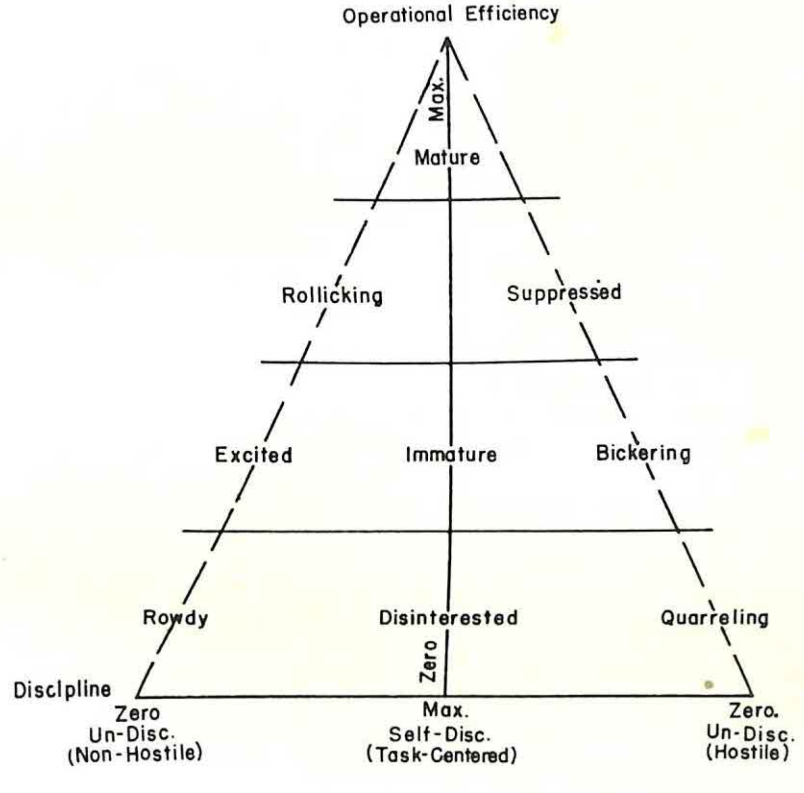


FIGURE 4

TYPES OF GROUPS WHICH OCCUR DURING THE OPERATIONS STAGE angle. Figure 4 shows the nine general group-types that have been observed. In this figure, groups falling on the left side of the triangle are non-hostile groups, those falling on the right side are hostile ones, those in the center are best described as task-centered. The lower a group's position on the triangle, the less efficient it is as an operating unit, and the further away it is from the center line, the less disciplined are its members. As in the Planning Stage, we again find that the amount and kind of self-discipline group members possess serves to limit their efficiency in accomplishing a particular task.

The nine types of groups shown in Figure 4 are briefly described as follows:

### Task-Centered Groups

1. The Mature Group—This group is characterized by a warm friendly, supportive atmosphere in which all members exhibit sustained interest in accomplishing the task rapidly and efficiently. Once children have performed jobs assigned them, they either socialize freely, but quietly, until the task is completed or pursue ongoing classroom activities. This group completes the task in approximately ten minutes or less. It is distinguished from all other groups by the fact that those children not directly concerned with the task of putting blocks together still remain involved in the work.

2. The Immature Group—The atmosphere of this group, although basically warm and friendly, is best described as "bumbling." Children exhibit a sincere concern for accomplishing the task, but their lack of technical skill in building an efficient plan of action renders their efforts ineffective. The majority mill around the problem table, putting blocks together in a more or less random fashion, trading blocks, taking them apart and reassembling them, usually to little avail. The warmth and friendliness of the atmosphere, however, prevents the group from breaking up despite their frustrating trial-and-error methods. This group rarely completes the problem in the allotted time, usually being less than three-fourths finished when the fifteen minutes are up.

3. The Disinterested Group—This group is characterized by extremely low involvement on the part of most members in the class. Because only a small group work on the problem, the working atmosphere is generally a quiet one and success, measured by the time necessary to complete the task, is high. Since the problem is no longer a group problem, the over-all rating on Operational Efficiency is near zero despite the high degree of objective achievement. This group is distinguished from the Rowdy and the Quarreling Groups by the fact that children who have withdrawn pursue acceptable (e.g., non-disruptive) classroom activities.

### Non-Hostile Groups

4. The Rollicking Groups—Characterized by an atmosphere which, although basically friendly, is loud, noisy and generally uproarious. Children not directly concerned with the building task crowd around the participants shouting directions, making joking comments about progress, and laughing loudly over any errors which are made. The melee generally becomes too much for some members, and they withdraw from it altogether—not in anger but for self-preservation. The uproar and lack of seriousness lowers this group's over-all efficiency.

5. The Excited Group—Children in this group are unable to control their excitement over the task. The atmosphere is charged with tenseness, voices are sharp and high, and movements are jerky as children work feverishly at the building task. In the excitement tempers become short and arguments frequent, but real hostility is absent as shown by the fact that outbursts are task-centered, rather than personal, and are short-lived. The noise, tenseness, and excitement prevent children from achieving a high success score, although they frequently manage to finish before the allotted time is up.

6. The Rowdy Group—This group is similar to the Excited Group, except that here all forms of control are so lacking that the excitement gets completely out-of-hand. The screaming, running, shouting, and yelling reach such a pitch that the examiner frequently is forced to stop the test. This group's efficiency is always near zero.

### Hostile Groups

7. The Suppressed Group—Characterized by an almost abnormal quiet. Children engaged in the building task speak in whispers or in a low murmur, and it is rarely possible to hear specific statements. Children not engaged in the building task sit quietly and idly in their seats doing nothing.

8. The Bickering Group—This group is characterized by an atmosphere of contained hostility. There is a continuous flow of negatively critical arguments about who should be at the problem table and how the puzzle ought to be built. The hostility apparent in these arguments is usually sufficient to cause many children to

withdraw from the field altogether with the attitude: "They think they're so smart—let them do it." Those who remain continue to bicker about how the task should be done; those who withdraw lose all interest in the work, sitting quietly (though often sullenly) at their desks or engaging in some type of play activity. Frequently, children hide their blocks so that the builders cannot complete their task.

9. The Quarreling Group—In this group the hostility gets completely out-of-hand. Quarreling over possession of blocks and fighting for positions at the problem table generally reach the point at which the test has to be stopped.

## Changes in Test Performance Over Problems

The present test, being concerned with the measurement of dynamic social and psychological forces which reveal themselves in changes in behavior rather than in consistency of behavior, does not possess reliability in the sense of having high inter-problem (e.g., inter-item) correlations. As a group progresses from the Planning Stage of problem one to the Operations Stage, and from there to the Planning Stage of problem two, etc., the nature of the psychological forces active in that particular group are revealed by the pattern of changes which occur in test behavior. It is hypothesized that if these forces are predominantly constructive in nature, the group will become more cohesive and more efficient over problems. Thus a group that scores Immature on problem one planning and Mature on problem one operations, generally scores Dependent or Semi-Controlled in subsequent planning sessions. If, on the other hand, these forces tend to be predominantly destructive in nature, the group will show a more or less rapid trend toward disintegration and inefficiency, the strength of these forces being revealed in the rapidity with which the group reaches the point of complete disintegration.

The explanation for these phenomena is based upon two assumptions. First, every group, regardless of the attributes or past experiences of its members, contains within it potentialities for both constructive and destructive group action. Second, the direction which these potentialities ultimately take is determined primarily by the experiences to which the members—as a group—have been subjected. Accordingly, groups in which constructive forces are stronger than destructive ones are hypothesized to be those in which children have been exposed to many happy and constructive working experiences in a group situation where respect for the abilities, needs, and feelings of others is the rule. Groups in which destructive forces are paramount are hypothesized to be ones in which primary emphasis has been placed upon individual work and individual accomplishment. Children in these groups lack the kind of experiences in which social relations skills may be learned, and thus fail in their efforts when suddenly placed in a situation which demands such skills. Groups in which destructive forces show themselves in hostile behaviors are believed to be ones in which children's experiences have been such as to generate marked feelings of rivalry, jealousy, bitterness, and frustration. Those in which destructive forces show themselves in rowdy, although non-hostile, behaviors are believed to be ones in which children have not been taught to handle their freedom wisely and constructively. There is no frustration, only a boredom which breaks into excited and welcomed horseplay when external controls are released.

These and similar hypotheses are as yet quite tentative, as are many of the findings here reported. Much research of a rigorous nature is needed before the test can be used with confidence and with accuracy. Research on reliability is now in progress, further use of the test hinging upon the outcome.

## Summary

This paper describes a situational test useful for evaluating the social relations skills of elementary school children in a group problem-solving situation. Data accumulated during the development of the instrument indicate that the observation and scoring procedures here reported possess satisfying reliability, but the rigorously designed study which will provide definitive answers to such questions has not been completed. The test is presented as a research technique which might prove valuable in studying the effects of various kinds of educational environments upon the behavior of children's groups.

## REFERENCES

1. Bales, Robert F. Interaction Process Analysis (Cambridge, Mass.: Addison-Wesley Press, 1951).
2. Kearney, Nolan C. Elementary School Objectives (New York: Russell Sage Foundation, 1953).
3. Krech, David and Crutchfield, Richard S. Theory and Problems of Social Psychology (New York: McGraw-Hill Book Co., 1948).
4. Lippitt, Ronald and White, Ralph K. "The 'Social Climate' of Children's Groups" in Child Behavior and Development, by Barker, Kounin, and Wright (New York: McGraw-Hill Book Co., 1943).
5. Pepinsky, H. B., and others. "The Criterion in Counseling: A Group Participation Scale," Journal of Abnormal and Social Psychology, XLVII (1952), pp. 415-19.

*Page 1*
Form 3526
Rev. 8–50

Read the instructions on page 2 of this form.

# STATEMENT REQUIRED BY THE ACT OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, AND JULY 2, 1946 (Title 39, United States Code, Section 233) SHOWING THE OWNERSHIP, MANAGEMENT, AND CIRCULATION OF

**Journal of Experimental Education** published **Quarterly**
(Insert exact title of publication) (State exact frequency of issue)

at **Madison, Wisconsin** for **October 1**, 19**59**
(Name of post office and State where publication has second-class entry)

1. The names and addresses of the publisher, editor, managing editor, and business managers are:

| | *Name* | *Address* |
|---|---|---|
| Publisher | Dembar Publications, Inc. | Madison, Wisconsin |
| Editor | A. S. Barr | Madison, Wisconsin |
| Managing editor | Jean W. Williams | Madison, Wisconsin |
| Business manager | Walter A. Frautschi | Madison, Wisconsin |

2. The owner is: (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual member, must be given.)

| *Name* | *Address* |
|---|---|
| Dembar Publications, Inc. | Madison, Wisconsin |
| A. S. Barr | Madison, Wisconsin |
| Walter A. Frautschi | Madison, Wisconsin |
| Dorothy J. Frautschi | Madison, Wisconsin |

3. The known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: **(If there are none, so state.)**

| *Name* | *Address* |
|---|---|
| | |
| | |
| | |
| | |

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required from daily, weekly, semiweekly, and triweekly newspapers only.)

Walter A. Frautschi
per John J. Frautschi, atty in fact
(Signature of editor, publisher, business manager, or owner)

Sworn to and subscribed before me this 21st day of October, 19 59

[SEAL]

Therese Yanna

(My commission expires September 30, 19 62)

16—18720-6

# The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1960 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 2, December 1959)

# DEVELOPMENT OF AN INSTRUMENT FOR STUDYING VERBAL BEHAVIORS IN A SECONDARY SCHOOL MATHEMATICS CLASSROOM

E. MURIEL J. WRIGHT
Washington University

MUCH OF the research relating to the improvement of classroom teaching has been dependent on measurements carried out in the pre-lesson and/or the post-lesson periods. This approach, valuable as it is, is indirect, and there is need for a more direct approach. It seems reasonable to suppose that direct study of the lesson by an observational technique would avoid certain of the variables encountered in the indirect approach and, at the same time, should possess intrinsic validity. Simultaneous consideration of the subject matter taught and the method of its development—two interdependent facets of the lesson—could also be achieved.

There follows, hereafter, a description of the design of an instrument for direct observation of the verbal interaction of teacher and pupils in a mathematics classroom. Support for this design is sought in a brief empirical study of aspects of validity and reliability of such an instrument including effective ways of reporting the observations.

## Design of the Instrument

Categories related to objectives of mathematics teaching formed the basis of the instrument. Trained observers, sampling the verbal interaction of teacher and pupils, classified behaviors by means of the defined categories. This permitted a quantitative record of the emphasis of the selected objectives.

### The Categories

Fairly general aims in the teaching of secondary school mathematics were selected to provide a basis for judgment of the developing lesson. These were sought in the literature dealing with general educational objectives, in the literature concerned particularly with mathematical objectives, in writings of specific mathematical interest and from personal teaching experience.

Categories for the classification of classroom behavior were selected from these statements of aims of mathematics teaching in the light of three criteria. Each aim was capable of careful definition, important for general and for mathematical education, and feasible of attainment in the secondary school. These categories fell naturally into three frames of reference: ability to think, appreciation of mathematics, and attitude of curiosity and initiative.

Frame A. Conscious developing (teacher) or use (teacher or pupil) of ability to think by, 1. analyzing, 2. synthesizing, 3. specializing, 4. generalizing.

Frame B. Conscious developing (teacher) or demonstration (teacher or pupil) of appreciation of mathematics—an evaluative approach to, 1. the methodology of mathematics, 2. the subject matter of mathematics, 3. the place of mathematics in other fields and areas, 4. the place of mathematics in history.

Frame C. Conscious fostering (teacher) or demonstration (teacher or pupil) of an attitude of curiosity and initiative, 1. enthusiasm for fresh knowledge, 2. independence of thought and action.

The requirements of an instrument for systematic observation (exclusiveness, exhaustiveness, and unity) had also to be met by this definition of categories. Logically and by practical exploratory use, these frames (and within each frame the several categories) were deemed to be exclusive. Provision of a neutral category for behaviors not falling in the defined frames ensured exhaustive classification. Finally, because the aims were defined to prove necessary and sufficient to encompass the aims of mathematics teaching, unity was assumed.

Detailed definition of the positive and negative aspects of each frame and of the several categories of each follows in Table I.

### The Classification of Behaviors in Terms of the Categories

Frames of Reference—These aims of ability to think, appreciation of mathematics, and an at-

TABLE I

DEFINITION OF THE SEVERAL CATEGORIES OF FRAMES OF REFERENCE, A, ABILITY TO THINK, B, APPRECIATION OF MATHEMATICS, C, AN ATTITUDE OF CURIOSITY AND INITIATIVE, AND THEIR POSITIVE AND NEGATIVE QUALITIES. PARTICULAR EXAMPLE AND SYMBOLIC CLASSIFICATION OF A BEHAVIOR IN EACH CATEGORY

| Frame | Quality | Category | Definition | Particular Examples of Behavior | Symbolic Classification |
|---|---|---|---|---|---|
| Ability to think, A | Behaviors which demonstrate clear and logical ability to think are classified as positive; while those showing confusion and illogic are classified as negative. | 1. Analyzing | Working from an assumption of of a desired conclusion towards accepted principles. | *(The problem: to show that the roots of a particular equation are real and unequal)<br>Pupil: If the roots of $16x^2-16x-51 = 0$ are real and unequal, then we should find the discriminant greater than zero. | $A_1+$ |
| | | 2. Synthesizing | Working from accepted principles.toward a desired conclusion. | Teacher: We have decided that the three shares of the whole dividend may be expressed as \$2x, \$3x, \$2x. We also know that this whole dividend amount to \$66. What may we conclude from these two statements? | $A_2+$ |
| | | 3. Specializing | The use of the significant attributes of a given set in an analogous set, or the applying of a given set to a smaller included set. | Pupil: This is a geometric series, so to sum "n" terms I'd use the formula $S = \frac{a(1-r^n)}{1-r}$ | $A_3+$ |
| | | 4. Generalizing | The recognition of significant attributes of a given set and the passing from the consideration of the given set to that of a larger inclusive set. | Pupil: That answer can't be right, it doesn't agree with the answer in back of the book. | $A_4-$ |
| Appreciation of Mathematics, B | Behaviors which demonstrate correct, concise or elegant examples of the dimensions are classified as positive; while those proving incorrect, | 1. Methodology | Awareness, recognition of and intellectual pleasure in understanding the characteristics of the field of mathematics in terms of its methodology. | Pupil: My answer is $x^2+y^2+2xy-4xy$. | $B_1-$ |

| | | | | | |
|---|---|---|---|---|---|
| | verbose, or inelegant are classified as negative. | | | | |
| | | 2. Subject matter | Awareness, recognition of and intellectual pleasure in understanding the characteristics of the field of mathematics in terms of its subject matter. | *(The problem: To show that the roots of a particular equation are real and unequal)<br>Pupil: If the roots of $16x^2-16x-51=0$ are real and unequal, then we should find the discriminant greater than zero. | $B_{2+}$ |
| | | 3. Other fields and areas | Awareness, recognition of and intellectual pleasure in understanding the place of mathematics in other fields and in applied areas. | Teacher: Do you recall a problem in your physics class where "cross-multiplication" can be used. | $B_{3+}$ |
| | | 4. Historical significance | Consideration of the place of mathematics in the historical development of civilization. | Pupil: I brought this article about computing machines. Part of it tells about different number systems: the one the Egyptians used, our decimal one, and another one for these machines. | $B_{4+}$ |
| Attitude of curiosity and initiative, C | Behaviors indicating definite presence of the dimensions related to the study of mathematics are classified as positive; while behaviors indicating relative absence of the dimensions are classified as negative. | 1. Curiosity | Enthusiasm for fresh knowledge. | Pupil: But that is a chemistry problem. | $C_{1-}$ |
| | | 2. Initiative | Independence of thought and action including perseverance and self-directed work towards the desired outcome. | *(The problem: To show that the roots of a particular equation are real and unequal)<br>Pupil: If the roots of $16x^2-16x-51=0$ are real and unequal, then we should find the discriminant greater than zero. | $C_{2+}$ |
| Neutral, C | No apparent effect on achievement of the defined aims. | | Behaviors that cannot be readily classified under at least two of the above frames A, B and C. | Teacher: Pass in your homework questions please.<br>Pupil: May I sharpen my pencil, Mr. Blake. | N |

*This illustrates that classification of each observed behavior takes place under the three defined frames of reference.

tude of curiosity and initiative constituted three frames of reference. Each verbal behavior occurring in the classroom was viewed in all three frames and was classified under one of the categories of each. This is shown in Table I. The first particular example of behavior is:

> "(The problem: to show that the roots of a particular equation are real and unequal) Pupil: 'If the roots of $16x^2 - 16x - 51 = 0$ are real and unequal, then we should find the discriminant greater than zero.' "

This was classified under Ability to Think as Analyzing, positive, or $A_1+$. In addition, this example occurs under Appreciation as Subject matter, positive, or $B_2+$, and under Attitude as Independence, positive, or $C_2+$. When not readily classifiable under at least two of the above frames a behavior was recorded as Neutral.

Single Behavior—The behavior observed was the verbal responses of teacher and pupils plus any concurrent blackboard development, vocal inflection or facial expression contributing to the meaning of this verbal interaction.

The single behavior was obtained by time sampling; the behavior of the first speaker in a fifteen-second interval. During each minute of observation two behaviors were classified. Of the four fifteen-second intervals occurring, the first and third were used for observation, the second and fourth for classification.

Arrington (1) indicated strongly that the role of time sampling in observational techniques depends on the ability, by its use, of obtaining more reliable information than would be be obtainable by other means. For this instrument two practical difficulties directed its use. First, it was found that many naturally complete behaviors, if artificially divided, might be classified in parts—losing, however, their contextual meaning. The close following of the individual phrase essential in the study of the interaction in the classroom might result readily in such sub-classification of complete behaviors. Second, because judgment and recording in three frames simultaneously was required, short periods had to be taken from the continuous interaction to complete this. Thus, omission of classification of variable amounts of the natural interaction would take place. Complete elimination of these difficulties would occur only if adequate recording devices, i.e., sound films, were practical. However, the two difficulties were diminished by time sampling: first, an interval of sufficient length to ensure classification of most behaviors was chosen; second, use of alternate intervals for classification permitted assessment of the amount of interaction unobserved.

Unit of Behavior—The single lesson provided a natural unit for observation of the classroom. Because of some variation in length of period from one class to another, it was decided to limit the observation to forty-five minutes. A total frequency of ninety behaviors (two per minute) was thus obtained for each standard unit of observation.

The observer viewed and classified behavior from a desk at the back of the classroom using duplicated recording sheets and a stop-watch.

### The Observers

The classification of behaviors under the defined categories was determined by the understanding, interpretation and inference of the observers. Care, therefore, was directed to adequate preparation in terms of mathematical understanding and of knowledge of the experimental procedures. The principal observer, 1, was the investigator holding degrees in mathematics and in education and with experience in the secondary classroom. The assistant observer, 2, was a fourth-year student at the University who was planning to become a teacher of mathematics in the secondary school. This student was majoring in mathematics and had completed most of her professional courses, but had no classroom teaching experience.

Experience and Training of the Observers—Observer 1's understanding of the problem had been developing during an eight-month period of experimental design. Her experience in recording was in the use of successively modified schedules and amounted to approximately sixty clock hours.

Observer 2 was trained as follows: General discussion of the purpose, scope and methods of the problem was followed by detailed study of the arbitrary definitions of behaviors as collected in an Obervers' Manual. Thereafter, three visits occurred for practice in classification of behaviors in the classroom. Beginning with five minute periods of classification in a single frame only, at the third visit the full 45-minute unit was studied in all three frames. Further experience was gained by repeated classification of filmed interaction, and by observation of fourteen units in the classroom itself.

Interaction Effects of the Observers and of His Subjects—Seeking to minimize the effect of the observer on the subject, Observer 1 met all of the principals and most of the teachers before the period of observation began, presenting the proposal and answering resultant questions. Detailed description of the observational device was purposely avoided. This was explained in terms of the possible behavioral effects and was accepted readily. On the first day of observation the observers were introduced to the pupils as visitors from Washington University interested in teaching. Probably because of the custom of students from the University visiting in these schools, the pupils generally did not show any particular interest in the observers.

Acknowledgement of the assistance given, and a brief outline of results obtained was sent to each teacher after the analysis was complete.

### The Classes

The twelve algebra classrooms in which the instrument was studied were in secondary schools in Metropolitan St. Louis in a setting of higher than average economic status. The classrooms were selected to be comparable in terms of size and organization of school, academic and professional qualifications of the teachers, broad admission requirements to high school, and heterogeneous grouping as far as mathematical ability was concerned.

Most of the lesson period was spent in group discussion conducted with the teacher at the blackboard, the pupils at their seats. There were occasional periods of supervised study.

## Testing of Validity

At this stage in the development of this instrument two types of validity—content and construct —and reliability were considered.

Assessment of content validity included study of the 15-second division at alternate intervals as a valid sample of the natural behaviors of a unit; also considered was the number of units needed to make adequate representation of the mean behaviors of a topic.

The study of construct validity was directed to assessment of logical relationship between behaviors revealed by this instrument and general teacher and pupil characteristics.

Reliability of the observer was assessed in two ways—agreement between observers, and agreement between repeated classifications of the same filmed interaction by the single observer.

### The 15-Second Division at Alternate Intervals

Comparisons of the chosen 15-second fixed division, and of a larger and shorter fixed division were made with a natural division of the interaction. It was essential that this be the first step in consideration of the design of the instrument because all observations in the classroom were to be made by this time sampling.

### Method

Four ten-minute films (4) were selected as approximating a class period. Typescripts of the sound tracks of these films were prepared. Artificial division of the typescripts into arbitrary time intervals of $7\frac{1}{2}$-second, 15-second, and 30-seconds from zero time was made. This was followed by a division by classifiable behaviors without regard to time. This latter division was termed the "natural" division, and a natural behavior was defined as "a continuous section of the defined interaction, chosen without regard to time, forming a behavior which, when considered in its context, is readily classibiable in A, B, or C, considered separately, or as N."

Then the behavior in each $7\frac{1}{2}$-second interval was classified by observers 1 and 2, individually, in each of the frames A, B, and C, or alternatively as N. Classification in each of the frames or as N was repeated for the 15-second and the 30-second intervals. It proved necessary to include an additional category "unclassifiable" because of the occurrence of a number of intervals, particularly in the $7\frac{1}{2}$-second division with too few words to permit classification.

Next, the typescripts were divided by observers 1 and 2, individually, by natural behaviors classifiable in frame A or alternatively as N. This was repeated for frame B and for frame C, lest the length of the natural interval in each frame did not coincide.

Points of disagreement in a classification for each division were resolved by discussion, providing a single joint classification by observers 1 and 2.

### Results

Totals of teacher and pupils behaviors were obtained for the four films taken as an entity. These totals for the $7\frac{1}{2}$-second, 15-second, the 30-second and the natural divisions were 324, 163, 82, and 175, respectively.

Ratios of total teacher-pupils behaviors for these same intervals were 2.59, 2.60, 2.77, and 2.02, respectively. In each fixed division, an enlarged teacher effect resulted.

Further consideration of the $7\frac{1}{2}$-second or the 30-second divisions was not warranted after noting wide disparity from the natural values in their total number of responses, the teacher-pupil ratio, and the lengths of the intervals.

A frequency distribution of the lengths of the natural division is illustrated in Figure 1. These lengths coincided in the classification of behaviors in each of the three frames.

Using the "t" test of significance of a single mean (3), probability levels of non-significant differences between totals of the 15-second division and of the natural division were obtained (5) and recorded in Table II. It was noted that in no frame was a significant difference apparent (significance level, $p = 0.05$).

Probability levels of the non-significant differences between totals of 15-second even and odd sets of intervals were recorded in Table III. It was noted that none of the frames showed significant differences at the significance level $p = 0.05$. Indeed, of the twelve comparisons made, only two probability levels less than 0.60 occurred.

The use of this "t" test necessitated the as-

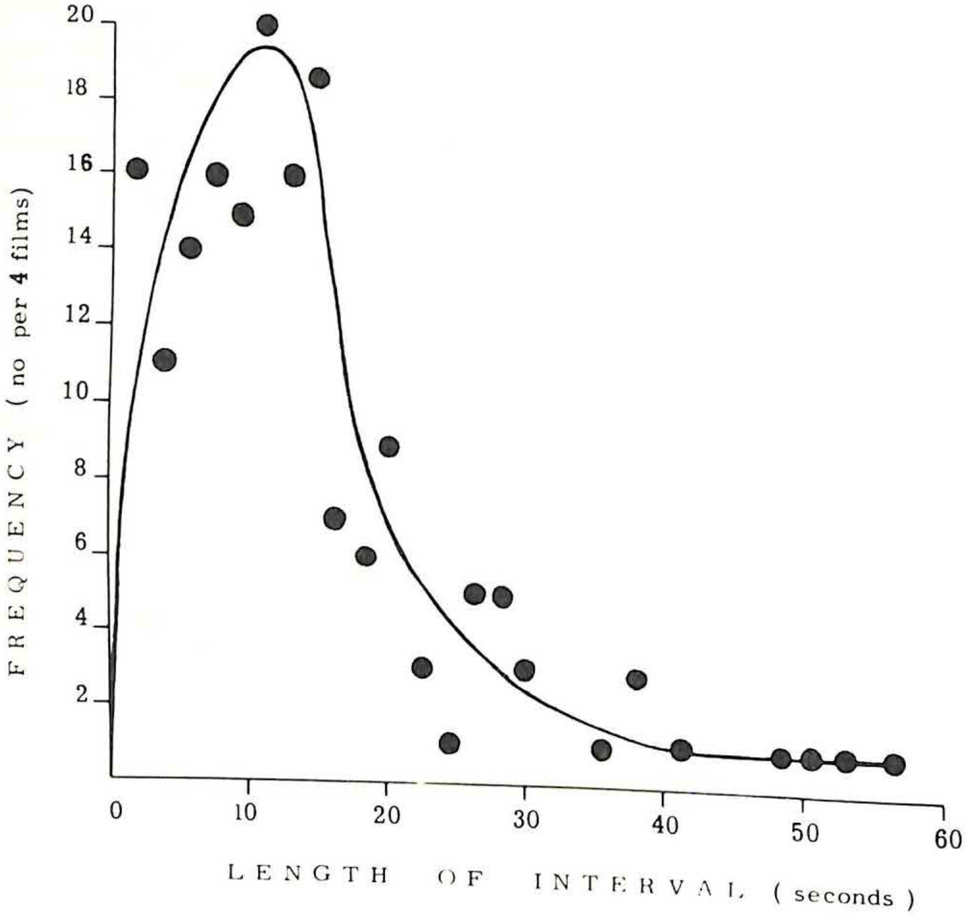


FIGURE 1. FREQUENCY DISTRIBUTION OF LENGTHS OF NATURAL INTERVALS FROM CLASSIFICATION OF TYPESCRIPTS OF SOUND TRACKS OF FOUR FILMS

## TABLE II

### PROBABILITY LEVELS BETWEEN 15-SECOND AND NATURAL DIVISIONS*

| Measure | Teacher | Pupil | Class |
|---|---|---|---|
| $P_A$ | 0.82 | 0.10 | 0.39 |
| $P_B$ | 0.35 | 0.20 | 0.45 |
| $P_C$ | 0.32 | 0.06 | 0.18 |
| $^{(A,B)}P_{(N,U)}$ | 0.33 | 0.70 | 0.75 |
| $^{(C)}P_{(N,U)}$ | 0.05 | 1.00 | 0.15 |

* Teacher, pupil and class behaviors in Ability to Think, A, Appreciation, B, and Attitude, C. Symbols for Neutral and Unclassifiable N, U. For A and B, n = 7; for C, n = 3; for N, U, n = 1. Use of "t" test of significance of a single mean. Significance level $p = 0.05$. Where $p > 0.05$, difference is not significant. Where $p < 0.05$, difference is significant.

## TABLE III

### PROBABILITY LEVELS OF SIGNIFICANT DIFFERENCES BETWEEN THE ODD AND EVEN SETS OF INTERVALS OF THE 15-SECOND DIVISION*

| Measure | Teacher | Pupil | Class |
|---|---|---|---|
| $P_A$ | 0.60 | 0.85 | 0.60 |
| $P_B$ | 0.09 | 0.85 | 0.13 |
| $P_C$ | 0.75 | 0.86 | 0.75 |
| $P_{(N, U)}$ | 1.00 | 1.00 | 1.00 |

* Teacher, pupil and class behaviors in Ability to Think, A, Appreciation, B, and Attitude, C. Symbols for Neutral and Unclassifiable N, U. For A and B, n = 7; for C, n = 3; for N, U, n = 1. Use of "t" test of significance of a single mean. Significance level $p = 0.05$. Where $p > 0.05$, difference is not significant. Where $p < 0.05$, difference is significant.

sumption of the continuity of the variable formed by the differences between the frequencies of corresponding categories in the fixed and in the natural divisions. Later this assumption was avoided by the study of each category separately.

### Discussion and Inferences

The $7\frac{1}{2}$-second division and the 30-second division were inappropriate as approximations to the natural division because of wide discrepancies apparent in total numbers of responses, teacher-pupil ratios of behaviors and lengths of intervals.

The 15-second division and the natural division compared closely in terms of total numbers of responses and lengths of intervals, while the ratio of teacher-pupil behaviors still indicated an enlarged teacher effect. However, no significant differences were shown to occur between corresponding teacher, pupils and class behaviors in frames A, B, or C, or Neutral and Unclassifiable. Further, no significant differences occurred between classifications of the even and odd alternate sets of 15-second intervals.

Recognizing the limitations of the use of any fixed division in approximating the observation of natural behaviors and in particular the enlarged teacher effect present in its teacher-pupils ratio, the 15-second division at alternate intervals was shown empirically to be a valid sample of the totality of behaviors of a unit.

Such comparison of the classifications of a fixed division and a natural division seems unusual in studies of validity of time sampling. For immediate description the criterion has been that of reliability. For conclusions on the general situation the accepted criterion has been that of the internal consistency of the data determined by comparison of different samples taken in the same manner in the same situation. It is noted that this assumes that variation other than that resulting from randomization does not occur between successive samples. In the classroom, while a general pattern typical of a teacher and his pupils may well emerge, some variation between successive lessons will be intentional or may result from such other factors as change in content.

## Number of Units Necessary to Represent Adequately the Mean Behavior of a Topic

Observation of classroom interaction using the 15-second sampling interval was then begun. It was necessary to determine empirically the minimum number of units of behavior which would provide representative estimates of mean behaviors of a topic.

### Method

Three classrooms studying the same topic in "Algebra 3" were visited by Observer 1. Classification of all units of behavior in that topic was carried out. The results were recorded as Classrooms I, II, and III.

Total of behaviors of each teacher and group of pupils were calculated for each category. For convenient comparison, these raw score frequencies were then adjusted proportionately in order that a standard total number of 100 responses was obtained for each class period.

Cumulative arithmetic averages, $x_i$, of adjusted frequencies of behaviors were calculated beginning with unit 2. Unit 1 was not considered in these because it was felt likely that the first lesson of a topic might vary markedly from the lessons of the succeeding days. The means for the complete topic, including unit 1, were defined as the theoretically best estimates for each category for the topic: $\overline{x}_i$.

Frequencies of individual categories were tested for the significance of their differences from the theoretical frequencies by considering the probability of occurrence of such differences in the binomial distribution.

Present in the use of this analysis was the assumption of independence of $x_i$ and $\overline{x}_i$. This assumption would prove justifiable in practice if the theoretical value were the mean of a very large number of units of behavior. The errors entailed by the assumption in this problem became quite marked as the number of units of behavior included in the cumulative averages increased towards the total number of units observed. It was felt, however, that consideration of individual categories by the binomial distribution included fewer assumptions than would the "t" tests as used above, or the $\chi^2$ test, which was possible in numerous investigations throughout the literature. In addition to the assumption present in this use of the binomial distribution, the use of $\chi^2$ in the present stage of analysis would include an assumption of independence of the several categories. While their exclusiveness has been argued logically and imposed by the method of classification, it was felt that empirical justification of an assumption of independence should precede use of $\chi^2$.

### Results

Probability levels of non-significant differences between successive cumulative averages and the theoretical averages were obtained for each category. A general gradual increase in the levels of probability occurred as the number of units included in the cumulative averages increased. In certain instances, this rise was discontinuous. In no category nor in any classroom were consistent sharp increases in the levels of probabilities observed.

Investigation followed of the differences between the four-day cumulative average (largest obtainable from all three classes) and the theoretical average in all three classes. For this average of four units the numbers of categories not significantly different from the theoretical values was noted for probability levels: $p = 0.70$, $p = 0.80$, and $p = 0.90$. These were approximately 4/5, 3/5 and 2/5, respectively.

Limited assessment followed of the consistency of the results of a single day with the mean scores. At a probability level of $p = 0.05$, less than 1/10 of the scores of individual categories for the single unit differed significantly from the mean category score.

### Discussion and Inferences

In seeking a practical limit to the number of days necessary for observation of a classroom, it has been decided that the totality of behaviors in a topic would be assumed to represent all behaviors occurring in the classroom. Cumulative averages of the units of a topic were then considered, in an attempt to determine a small number of units that could be used to approximate the topic. It was considered that if such a small number of units existed within a topic, that a sharp rise in the probability of non-significant differences should occur throughout all the categories and in each experiment on a particular day. Such a sharp rise did not occur. The steady though occasionally discontinuous rise which was met could well be accounted for by the increasing degree of dependency between cumulative averages and theoretical values as numbers of units increased. It was felt, therefore, that the small number of units providing the theoretical values of the categories prevented an answer to this problem. A profitable study might well result from consideration of the same problem using observations collected over a very large number of units, eliminating the assumption concerning the topic behaviors equalling the totality of all classroom behaviors, and minimizing the dependence of cumulative averages on the mean of all units.

It was necessary, practically, for the remaining observations in this series to limit the number of units of observation in each classroom. It was decided to visit classrooms during four units of behavior, and presuming that the first and last days of a topic might well be atypical, to collect data only on intervening days of topic.

## The Level of Agreement of Observers in Applying the Arbitrary Definitions of Classification of Behaviors

As a prerequisite to validity, it was also necessary to assess the arbitrary definitions of classification of behaviors in the light of observer agreement.

The test of observer reliability was twofold: first, the level of agreement in classification achieved between observers in single units of classroom behaviors, and second, the level of agreement in classification reached between repeated observations by the same observer of a single unit of behavior.

### Method

Two observers took part in the complete set of observations. Fourteen class periods were classified by both Observers 1 and 2. In addition each observer made repeated classification of a single recorded unit of behavior at three weekly intervals throughout the experimental period.

### Results

Level of Agreement in Classification Between Observers—Scattergrams of the corresponding scores of Observers 1 and 2 in all categories of teacher, pupil and class behaviors for units of observation 1, 7 and 14 are shown in Figure 2. By day 14, the results indicated a marked rise in observer agreement over those for units 1 and 7, and a high absolute achievement of agreement in the close clustering of points about the line of perfect agreement.

The probability levels of individual category frequencies of behaviors classified by Observer 2 being different from those as classified by Observer 1 were recorded. A summary of the number of differences non-significant at the probability level $p = 0.70$, was illustrated graphically on a linear scale in Figure 3.

After the first four observations a marked trend towards increasing agreement was noted.

Level of Agreement in Classification Between Repeated Observations of a Single Unit by the Same Observer—Observed frequencies and deviations of these from "true" scores in each category were recorded for Observer 1 and for Observer 2. Illustration was made of deviation from true values in Figures 4 and 5 for Observers 1 and 2 respectively. The horizontal axis represented zero deviation from the theoretical frequency.

These figures were used to compare observed with theoretical frequencies; successive observations by the same observer with each other; teacher, pupils and class patterns; and finally, the classifications of the two observers.

### Discussion and Inferences

Realization of the lack of a well-developed methodology of observational techniques cited by Heyns and Lippitt (7) continued to grow in seeking to assess observer agreement.

Use of means of observer scores in statistical

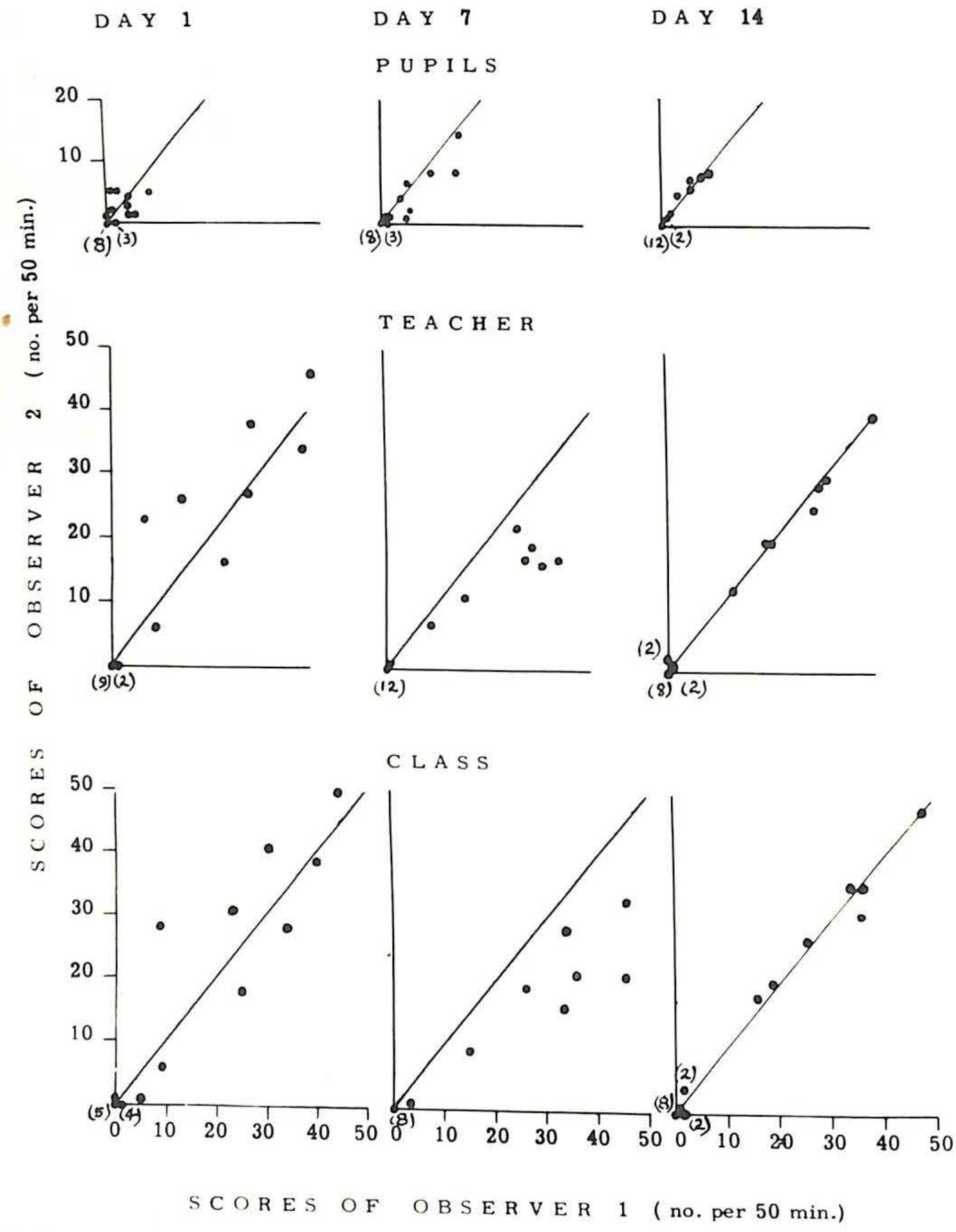


FIGURE 2. SCATTERGRAM COMPARISON OF SCORES IN CLASSIFICATIONS OF OBSERVERS 1 AND 2 AT THREE STAGES OF THE TRAINING PERIOD FOR OBSERVER 2: DAY 1, DAY 7, DAY 14. SCORES COMPARED IN THE 21 CATEGORIES OF A, B, C AND N. TOTALS OF SCORES IN A + N = B + N = C + N = 100. SOME MASKING OF REPEATED VALUES OCCURS (See bracketed numbers of categories.)

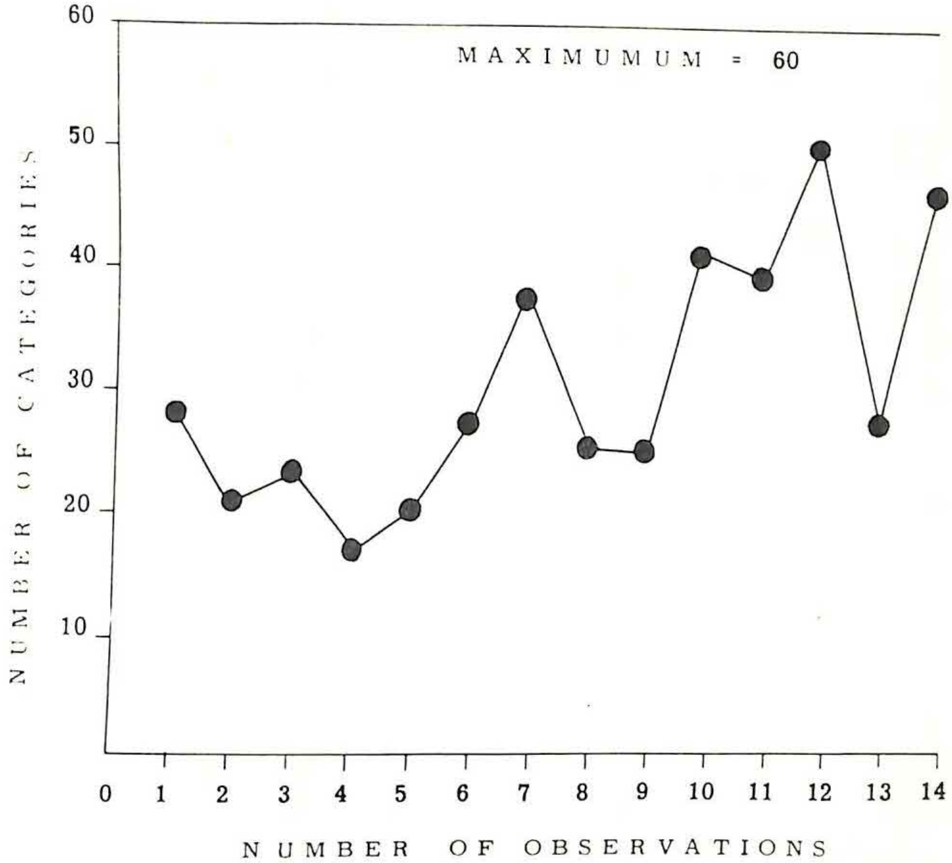


FIGURE 3. NUMBER OF CATEGORIES WITH NON-SIGNIFICANT DIFFERENCES IN CLASSIFICATION BY OBSERVERS 1 AND 2 FOR EACH OF FOURTEEN UNITS OF CLASSROOM OBSERVATION. USE OF BINOMIAL DISTRIBUTION LEVELS OF SIGNIFICANCE: $p = 0.70$: NON-SIGNIFICANT DIFFERENCES WHERE PROBABILITY LEVEL $\geq 0.70$

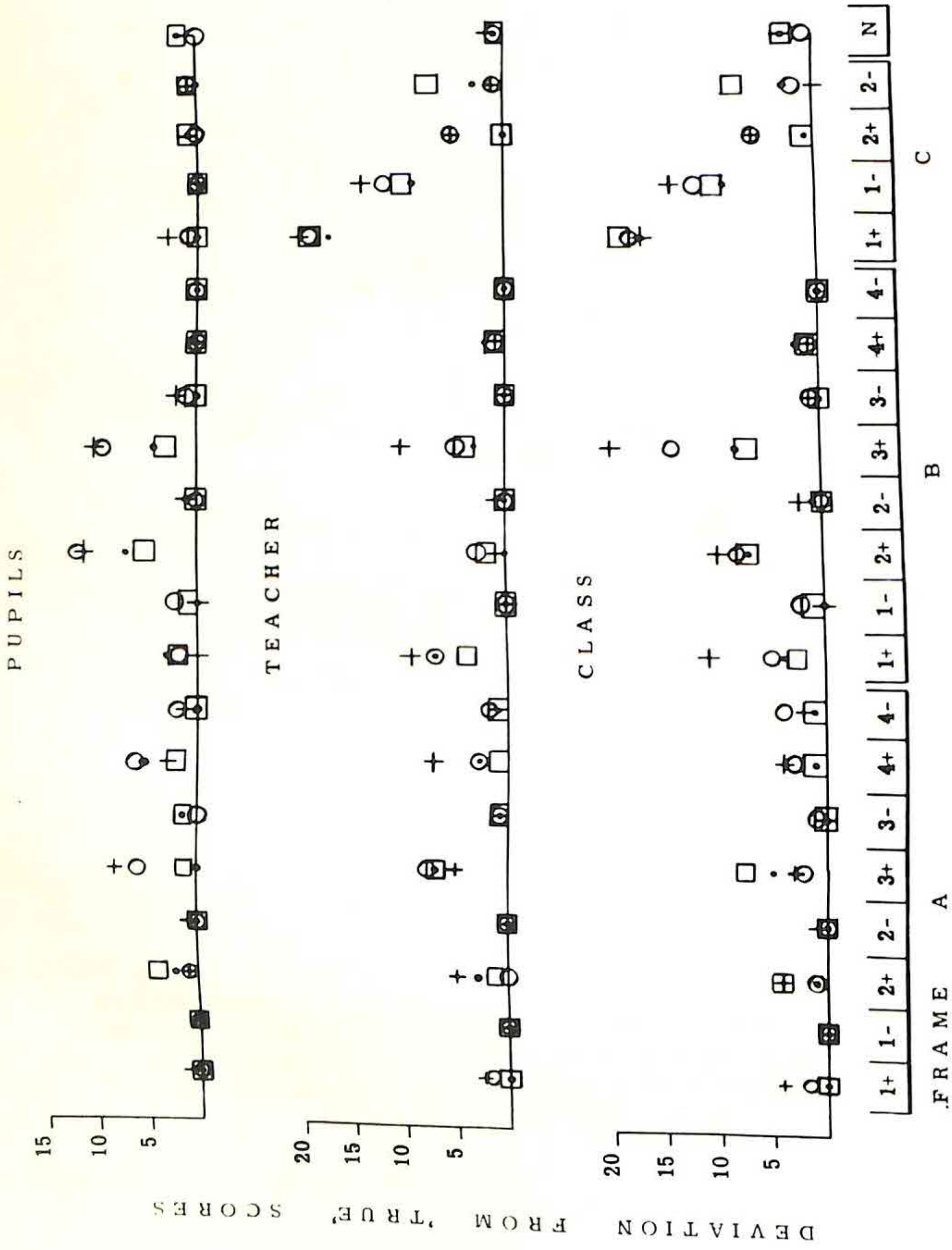


FIGURE 4. DEVIATION OF INDIVIDUAL CATEGORY FREQUENCIES FROM 'TRUE' SCORES IN CLASSIFICATIONS 1, 2, 3 AND 4: X, O, ., AND □ RESPECTIVELY. REPEAT CLASSIFICATION BY OBSERVER 1 OF THE FOUR FILMS

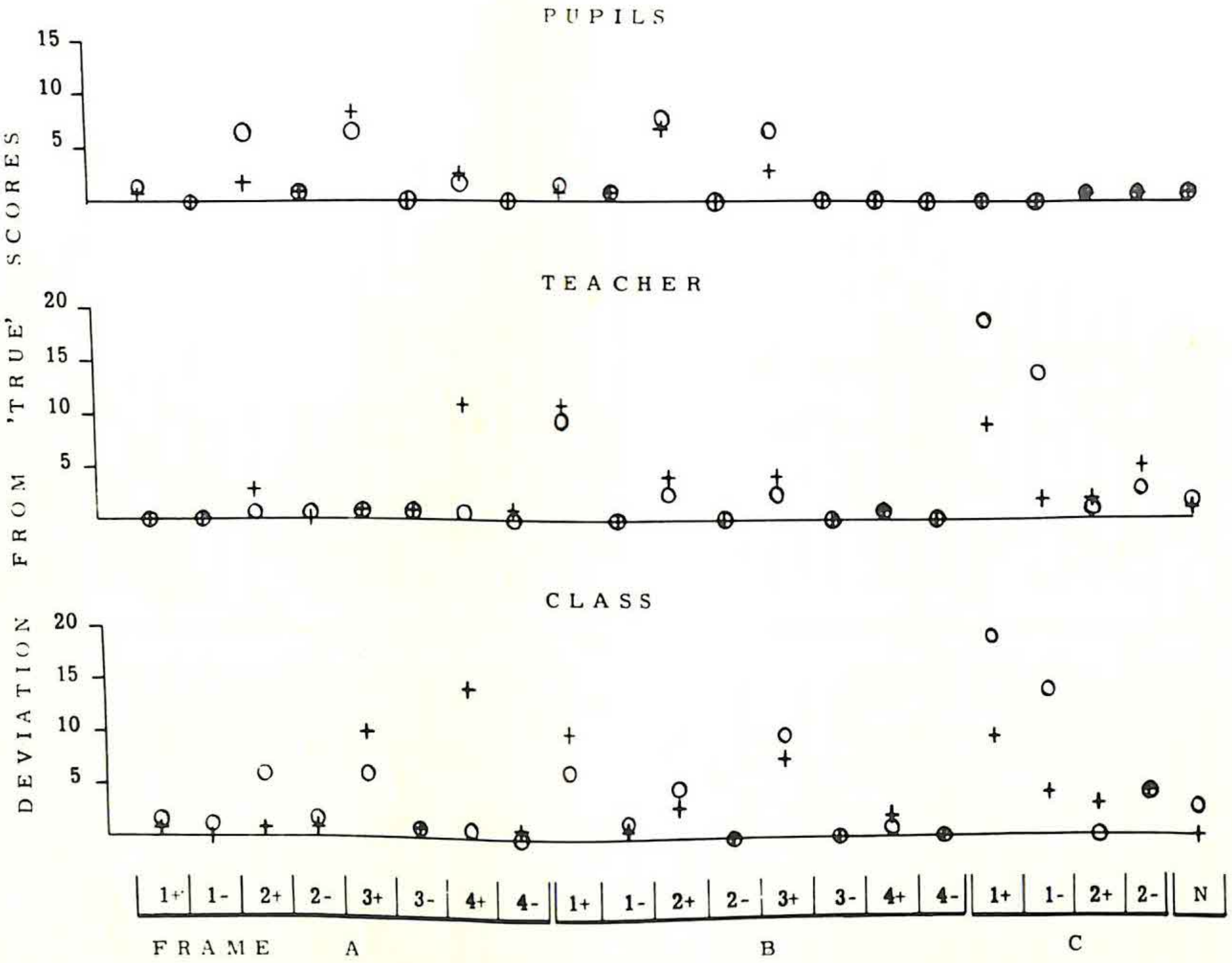


FIGURE 5. DEVIATION OF INDIVIDUAL CATEGORY FREQUENCIES FROM 'TRUE' SCORES IN CLASSIFICATIONS 1 AND 2: X AND O RESPECTIVELY.

assessment of between observer agreement was recommended by such investigators as Bales (2) and Steinzor (8) without indication of the dependence of this mean on the individual observer scores compared with it. In addition, the level of experience of the two observers was not comparable, preventing use of the technique suggested by Guetzkow (6). Differences between the scores of Observers 1 and 2 were consequently studied by considering the probability of Observer 2's scores being different from theoretical (Observer 1's) scores, by means of the binomial distribution.

In the area of level of agreement between repeated observations by the same observer of a single unit of behavior, the probability of differences occurring between observed and theoretical results also was studied by means of the binomial distribution. However, the nature of the scores themselves induced a further difficulty. Where frequencies in individual categories were low, and in particular where the theoretical frequency of a category equalled zero, the resultant 0.00 probability presented a markedly distorted picture. Therefore, it was decided to indicate deviation from the theoretical results at this stage by absolute values, and to illustrate this on a linear scale in Figures 4 and 5.

Level of Agreement in Classification Between Observers—Agreement between the two observers increased rapidly during the 14 units of observation until 50/60 categories were not significantly different at a probability level p = 0.70. Unfortunately, no evidence of use of the binomial expansion in a similar type of test was found in the literature to enable direct comparison with other studies. However, the level of agreement reached seemed to be entering a favorable level.

Because only two observers were compared, no conclusion may be drawn beyond the immediate situation. However, the results would infer success of the training methods for Observer 2. In addition, the level of agreement achieved allowed inference of the validity of the classification of the behaviors observed in this study.

Level of Agreement in Classification Reached Between Repeated Observations by the Same Observer of a Single Unit of Behavior—Because of overall small deviation from theoretical values, and because of consistency in this deviation between successive trials in the results of Observer 1 throughout the observational period (which began after the 2nd trial recorded here) was fairly consistent. It may also be inferred from the gradually decreasing values of the deviations in successive trials that some increased skill with additional experience was gained. Similar considerations follow from the two trials of Observer 2.

Two categories showed marked deviation from the theoretical values, despite consistent scores in successive trials and between observers: the positive and negative aspects of curiosity. This effect had not been unexpected. Discussion of interpretation of behaviors in this frame of attitude had brought out the importance of inflection of voice, and the effect of overall lesson development. It had been considered that these two factors might well be hidden or minimized in the close examination of a typescript as used in this part of the study. This emphasized the importance of direct classroom observation.

## Magnitudes and Variability of Mean Behaviors of Teachers and Pupils

These considerations of the validity of the sampling interval, the minimum number of units to be observed, and the level of agreement of the observers necessarily preceded interpretation of results.

Three main questions regarding classroom behavior observed were raised. First, were the mean behaviors in Algebra 1 classes significantly different from those of Algebra 3? Second, how should the magnitudes and variability of the selected aims be combined in describing the classroom? Finally, were known characteristics of teacher or pupils logically related to significant differences of behavior revealed by the instrument?

## Method

Materials Used—Interaction in six classrooms of each of Algebra 1 and Algebra 3 was classified by Observer 1 for four units of behavior.

The "t" test of significance of the difference between means was applied to corresponding behaviors in Algebra 1 and 3 ($p = 0.05$) (ref. 3).

Scores of individual categories of teacher and pupils were illustrated in two wyas: in absolute value on a bar graph, and in relation to mean values, by a profile comparison. In preparing the profiles, it was noted that scores of negative categories were inherently negative. Thus the mean - 1 p. e. was greater in absolute value than the mean. It was also felt that while a certain number of neutral behaviors was probably necessary for the conduct of a lesson, an incidence much higher than the mean value would suggest poor class management. Therefore, as in the negative categories, the scores of the neutral category was considered negative. However, since it seemed likely that some neutral behaviors were necessary, caution was marked in interpreting values falling far to the right as indicating high performance.

The scores of several of the twelve classes were then examined for differences in the observed behaviors that might logically be expected from certain known characteristics. This is exampled by Class XI. In Class XI, current extra-curricular responsibilities had limited the preparation time of the teacher. It was felt that this might be evidenced in class by an increased number of teacher behaviors. In particular, actual planning of the problems to be used in the lesson period itself would include a greater need for teacher demonstration of analyzing and/or synthesizing rather encouragement of this in the pupils. It was also presumed that encouragement of initiative would be low to avoid revelation of inadequate preparation. the scores in pupil initiative might be expected to indicate whether this lack of teacher preparation was a common experience: relatively high $C_{2-}$ would be likely to endorse this point of view. In addition, it was expected that pupil interaction would tend to contribute to the solution rather than the understanding of the problem—thus an emphasis in pupil behavior of analyzing and/or synthesizing and a decrease in specializing. Therefore, comparison of the results of Class XI with means of Algebra 1 was made in the following areas:

1. Totals of teacher and pupil behaviors
2. Teacher behaviors in $A_{2+}$, $A_{3+}$, $C_{2-}$, and $C_{2-}$
3. Pupil behaviors in $A_{2+}$, $A_{3+}$, $C_{2+}$, and $C_{2-}$

## Results

Means and standard deviations of raw scores forming representative behaviors were obtained for each classroom. From these means and standard deviations for the Algebra 1 and the Algebra 3 classrooms were calculated. Using the "t" test of significance of difference between means, and a probability level of 0.05, only two of the twenty-one categories of teacher behaviors were found to be different. These were positive analyzing and positive synthesizing. In each case the mean frequency of the teachers in Algebra 1 was less than the mean frequency of the teachers in Algebra 3. One category of pupil behavior was different. This was negative curiosity, in which mean Algebra 1 was greater than mean Algebra 3.

A bar graph illustrating the mean behavior of teachers and pupils in Algebra 1 is shown in Figure 6. Most of the behaviors in each were related to the three frames of reference, only thirteen in one hundred proving neutral in these areas. Consideration of the number of behaviors recorded for teacher and pupils separately emphasized the major part the teacher played in class interaction. Despite one teacher and approximately 25 pupils making up a class, five teacher behaviors occurred to every two pupil behaviors.

Using the profile of Class XI for comparison with mean values, the following differences were assessed: (1) Totals of teacher and pupil behaviors of Class XI were tested for their difference from mean behaviors of six classrooms using $\chi^2$. Corrected for continuity, $\chi^2 = 1.18$. This was less than that of $\chi^2$ at the chosen significance level, viz., 3.84, so goodness of fit was established. Thus the hypothesis that the number of teacher behaviors would be significantly less than mean teacher behaviors was not supported. (2) From the profile comparison, Figure 7, it was noted that teacher frequency in positive synthesizing, $A_{2+}$, was high, teacher frequency in positive specializing, $A_{3+}$ was low, teacher frequency in positive independence, $C_{2+}$, was low, and in negative independence, $C_{2-}$, was average. The first three relationships with the mean were thus shown to agree with the earlier hypothesis. (3) From the profile comparison, Figure 7, it was noted that pupil frequency in positive synthesizing, $A_{2+}$, was high, pupil frequency in positive specializing, $A_{3+}$ was low, pupil frequency in positive independence, $C_{2+}$, was low, and in negative independence, $C_{2-}$, was low. The first three relationships agree with the hypothesis. The point of view that the type of lessons observed occurred regularly was not supported.

## Discussion and Inferences

Generally, mean behaviors of teacher, pupils or class in Algebra 1 were not different to those shown in Algebra 3. It may be inferred that the two samples studied were drawn from the same population. Thus in this instance, differences in specific subject matter or in age of pupils did not affect significantly the pattern of behaviors. This was not surprising in terms of subject matter for the courses represented two parts of the overall algebra program for the state with differences in content rather than in difficulty of subject matter. It was possible, however, that the differing ages of the pupils in the two samples, and the probable selection of pupils preceding the course in Algebra 3 might have resulted in general significant differences in pattern. Later study, then, might be carried out within the secondary school algebra program with less regard for the formal divisions occurring in the schools.

The order of emphasis of categories in the aim ability to think was as follows: specializing, synthesizing and generalizing (approximately equal), analyzing. It was felt that the order of the first three was reasonable, but that the small incidence of behaviors in analyzing was undesirable in the light of the widespread usefulness of this mode of thought. The order of emphasis in the aim of appreciation was as follows: methodology and subject matter (approximately equal), other fields and areas, historical development (no behaviors). It was felt that emphasis of the first two was most desirable in mastery of a body of knowledge, but that greater emphasis in the latter two would contribute to the pupils' enthusiasm and their understanding.

The order of emphasis in attitude was negative

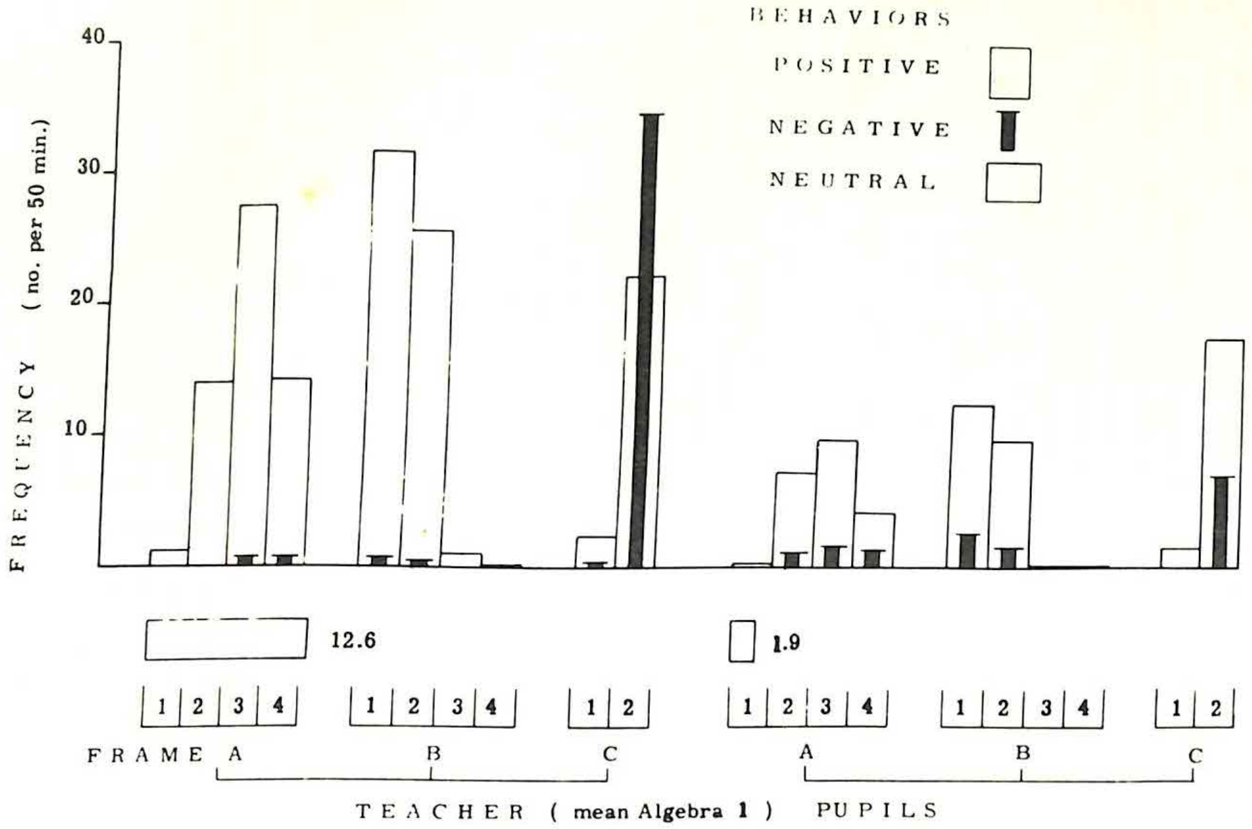


FIGURE 6. FREQUENCY (NUMBER PER 50 MINUTES) OF MEAN BEHAVIORS OF TEACHERS AND PUPILS, SEPARATELY, OF SIX CLASSES OF ALGEBRA 1: POSITIVE AND NEGATIVE BEHAVIORS IN EACH CATEGORY OF FRAMES ABILITY TO THINK, A, APPRECIATION, B, AND ATTITUDE, C, OR NEUTRAL

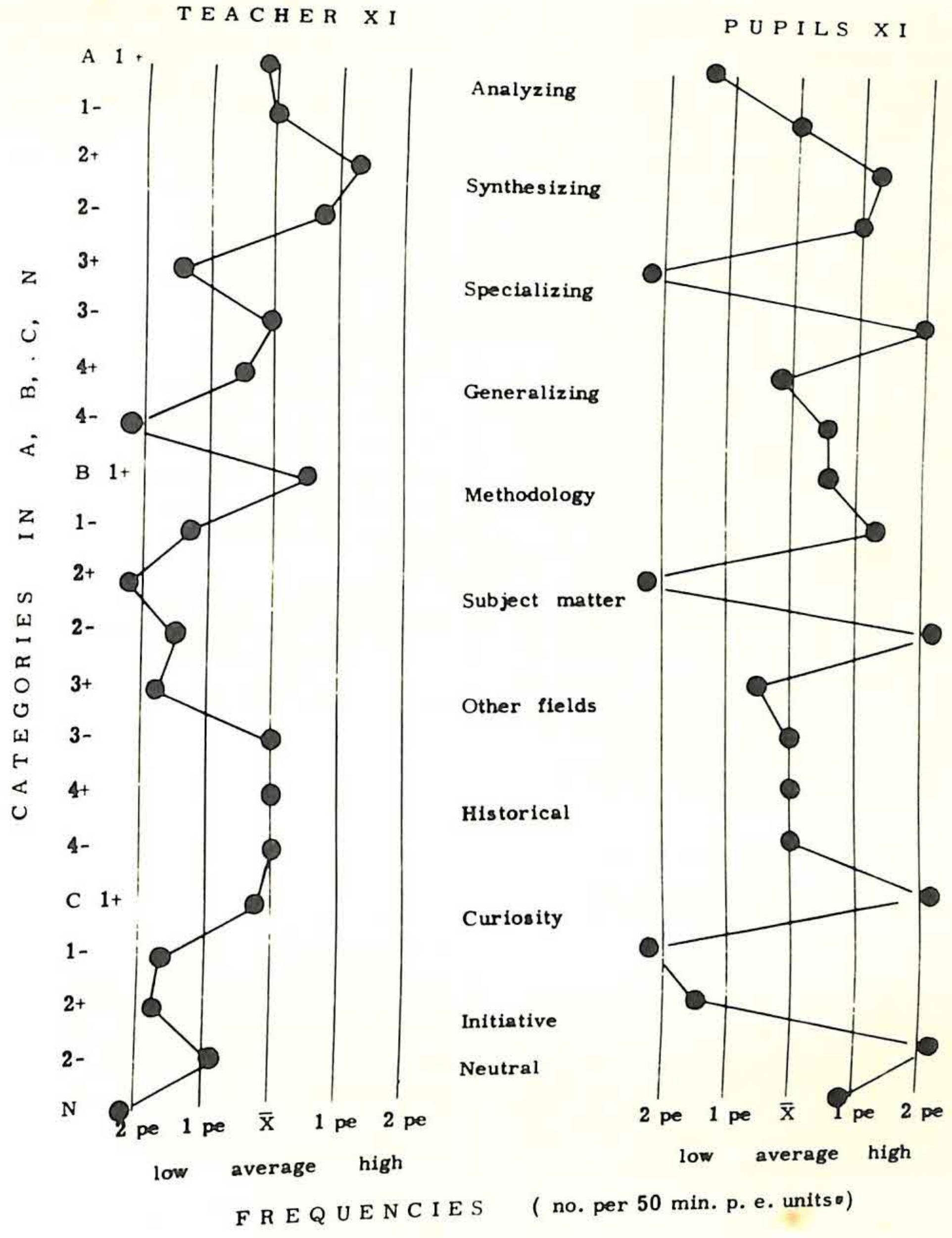


FIGURE 7. PROFILE COMPARISON OF THE FREQUENCIES IN EACH CATEGORY OF A, B, AND C, OR N OF TEACHER XI AND PUPILS XI WITH MEAN VALUES OF THE SIX CLASSES IN ALGEBRA 1

and positive independence (approximately equal), and curiosity (negligible). The low incidence of curiosity was disturbing in considering the potential force for leaning unexploited. Greater emphasis of positive independence was also felt desirable, although as the negative aspect included an absence of encouragement as well as active discouragement uncertainty as to the level to be attained was expressed.

Examination of the profile comparison in light of known characteristics and expected differences supported the validity of the instrument, for generally, expected differences did occur.

Development of these two aspects of examination of scores—in absolute value and in profile comparison—would seem promising in seeking a means of evaluating the performance with this instrument.

## Summary

The design of an instrument for direct study of the secondary school mathematics classroom is described. This design required the selection of criteria for judgment in secondary school mathematics; the development of an observational schedule based on these criteria; and the empirical assessment of the instrument in the classroom.

As criteria for judgment, careful definition was made of certain aims of mathematics teaching. These aims were defined in three frames of reference each of several categories:

A—Ability to think, analyzing, synthesizing, specializing, generalizing;
B—Appreciation of mathematics, methodology, subject matter, other fields and areas, historical significance;
C—Attitude of curiosity and initiative, enthusiasm for fresh knowledge, and independence.

Single behaviors in the classroom were classified as positive or negative achievement of teacher or pupils in all three of these frames of reference, or as neutral. The single act was the predominant behavior of the first speaker in a fifteen-second sampling interval. Alternate inter- were observed, recording being carried out in the intervening intervals. The period of observation was forty-five minutes in length, resulting in the classification of ninety behaviors. Each classroom was visited for four such periods, the mean behaviors resulting in "representative" behaviors for that classroom.

The observers were the investigator, and a fourth-year student, a potential teacher of mathematics. It was essential that the observer be skilled in subject matter as well as in the techniques of the observation schedule.

Empirical assessment in algebra classrooms followed this period of design of the instrument. Analysis of the results permitted the following inferences:

By an unusual direct comparison of behaviors yielded by time sampling and by a defined natural division of the verbal interaction, the choice of the 15-second interval was supported. In addition, a limited consideration of the consistency from day to day of observations by this sampling interval showed a reasonable level of agreement. Finally, the facility of its use in the classroom was so marked, that the validity of this sampling interval was felt to be established.

The second sampling procedure used was that of observing each classroom for a limited number of days. No conclusive answer to this problem was obtained. Practically, it was decided that for the rest of the present study the sample should consist of four days of observation of each classroom.

A high level of consistency in interpretation of behaviors was shown by the principal observer, and to a lesser degree by the assistant observer who was still gaining observational experience. These findings supported the classification of behaviors determined.

The use of bar graphs of absolute values of teacher and pupil behaviors, and profile comparisons of individual with mean values permitted detailed description of actual emphasis of aims in the classroom and comparison of individual classrooms with mean values. Consideration of certain known characteristics of teachers or pupils which were in agreement with differences shown by the instrument supported the validity of the descriptions yielded. Such profile comparison may well prove the basis for development of the instrument for evaluative purposes.

## General Appraisal of the Instrument

Because the basis of judgment for this instrument is that of aims of mathematics teaching, the instrument is theoretically limited to use in mathematics classrooms. However, the surprisingly general nature of the statement of aims might well bear consideration for application by specialists in other subject matter fields to their own situation. If such application was successful, general systematic comparison of teacher or pupil behaviors, independent of subject matter yet including in the observer interpretation the quality of the subject matter field, could be attained.

It is noted that the observers must be experienced in the subject matter field as well as trained in the particular observational procedures.

The behaviors classified with this instrument are those of verbal interaction or its attendant demonstration. Thus the description of the lesson period yielded by the instrument is limited to the amount of verbal class interaction occurring. In classrooms where long periods of supervised study are regularly used, this limitation may be so great that a picture of the in-lesson period by this instrument would be impossible.

The recording of class interaction in the present study has been divided into teacher scores and pupil scores. The pupils' scores are recorded as a group, thus no indication is available of individual pupil performance. In addition, discussion between teacher and individual pupils outside of the class interaction period—for instance while some of the class is studying—has not been considered. Thus a possible area of great teacher and pupil variation is untapped by the present observational procedure.

Currently another limitation exists in the need for preparation of the observer by the investigator. This will be necessary until further study of the training procedures can be carried out.

## REFERENCES

1. Arrington, Ruth E. "Time Sampling in Studies of Social Behavior," Psychological Bulletin, XL (1943), pp. 81-121.

2. Bales, R. F. Interaction Process Analysis (Cambridge, Mass.: Addison-Wesley, 1954).

3. Chambers, E. J. Statistical Calculation (London: Cambridge University Press, 1946).

4. Films: "$(A + B)^2$, International Film Bureau, 1954, Educational Film Guide, 1954-1958, p. 1.
   "How to Find the Answer" Coronet Films, 1949, p. 440.
   "Language of Mathematics" Coronet Films, 1950, p. 440.
   "Parallel Lines" Johnson Hunt, 1951, p. 444
   All in Educational Film Guide, 1953.

5. Fisher, R. J., and Yates, F. Statistical Tables for Biological, Agricultural and Medical Research (London and Edinburgh: Oliver and Boyd, 1953).

6. Guetzkow, H. S. "Unitizing and Categorizing Problems in Coding Qualitative Data," Journal of Clinical Psychology, VI (1950), pp. 47-58.

7. Heyns, R., and Lippitt, T. "Systematic Observational Techniques," in Handbook of Social Psychology (Cambridge, Mass: Addison-Wesley, 1954).

8. Steinzor, B. "The Development and Evaluation of a Measure of Social Interaction," Human Relations, II (1949), pp. 103-122.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 2, December 1959)

# THE SEMANTIC METHOD FOR TESTING VOCABULARY

MILDRED BERWICK
Boston College
Chestnut Hill, Massachusetts

## Purpose of this Study

THE BASIC objective of this study was the development of vocabulary tests that adequately measure knowledge of multi-meaning words by fourth-grade pupils through eighth grade. It was believed that this objective could be reached most effectively by:

1. Constructing vocabulary tests involving selections of meaning in relation to a pupil's interpretation of a word symbol.
2. Measuring more meanings per word than previous vocabulary tests. To illustrate the single meaning test, an item taken from the Metropolitan Tests, Elementary Reading Test, Form R, is shown below:

9. sack seat hold bag lift box

The Multi-Meaning Vocabulary Tests, in the item for the word "sack," include several meanings of the word. To illustrate:

sack
(x) a large bag
( ) to hold
(x) a loose coat
(x) to discharge
( ) some ashes
(x) to plunder
(x) a white wine
( ) to hunt

3. Measuring more meanings in approximately the same time limits. Individuals were able to designate all the meanings of a word in about the same time limits as they would need for finding the single meaning response.
4. Determining grade levels of meaning for each multi-meaning word used in the test for grades four through grade eight. No claim is made that this is a classification of multi-meaning words according to grade levels, but pupils' performance on each item demonstrated their proficiency or failure in identifying the meaning at each grade level included in the test.

## Development of the Test

### Method of Selecting Words for the Test

The Teacher's Word Book of 30,000 words was used as a source for securing a basic vocabulary list. The words in this book are not classified according to the grade for which they are intended, but are arranged in alphabetical order and coded to frequency of occurrence; therefore, each of the 30,000 words was examined, and all the words beginning with the first grade and ending with the eighth grade were reclassified according to grade levels. The number of words obtained for each grade is shown in Figure 1 and also the total number of words.

It will be noted that some lists were combined, having one list for two grades while other grades had their own lists. This grouping was used in the Teacher's Word Book of 30,000 words.

Since it was desired that the final selection of words for the test was the best sampling of multi-meaning words possible, several types of discarding were done and criteria to be met were established from the large initial list of 8008 words.

The first discarding was:

1. All prepositions, conjunctions, adverbs, personal pronouns, past participles, plurals and proper names were eliminated. The number of words retained after this first discard is shown in Figure 2. The results of the first discard reduced the initial list to 6504.

The next discarding was:

2. To discard all single meaning words. Each of the 6504 words was looked up in the Second Unabridged Edition of Webster's New International Dictionary, and if only one meaning was listed for a word, it was eliminated from the list.

Inspection of this graph shows that the list was reduced 612 words by this discard.

Since it was considered basic to the purpose of the test to measure at least three meanings of a word, the requirement was set up that a word was to have more than five dictionary meanings to re-

FIGURE 1

NUMBEF OF WORDS LISTED FOR EACH GRADE FROM THE THORNDIKE BASIC VOCABULARY LIST FOR GRADES 1-8

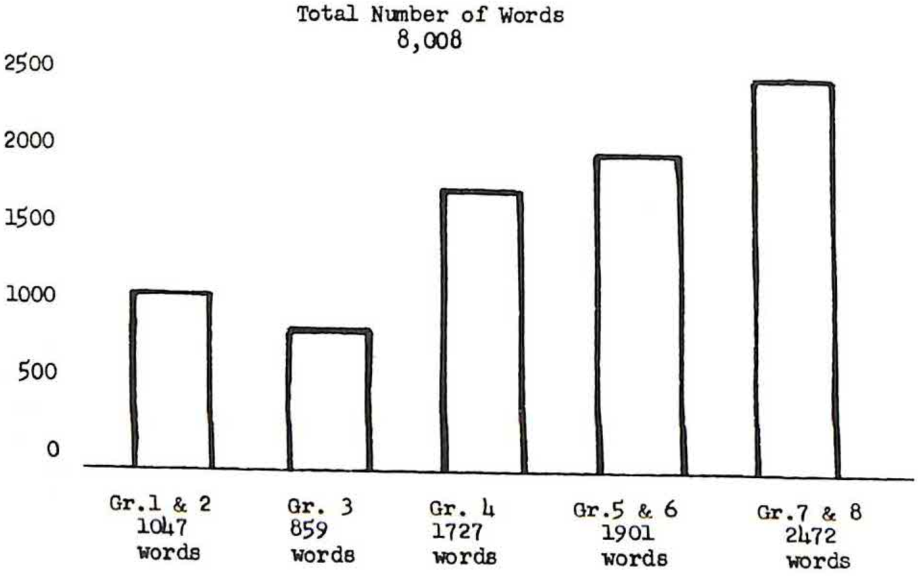


FIGURE 2

NUMBER OF WORDS LISTED FOR EACH GRADE AFTER CERTAIN PARTS OF SPEECH WERE ELIMINATED

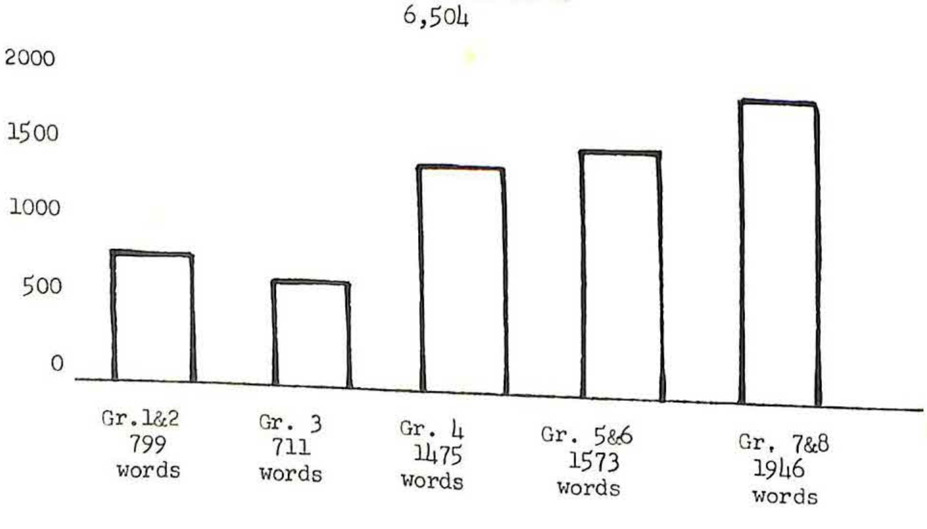


FIGURE 3

NUMBER OF WORDS LISTED FOR EACH GRADE AFTER THE SINGLE MEANING WORDS WERE ELIMINATED

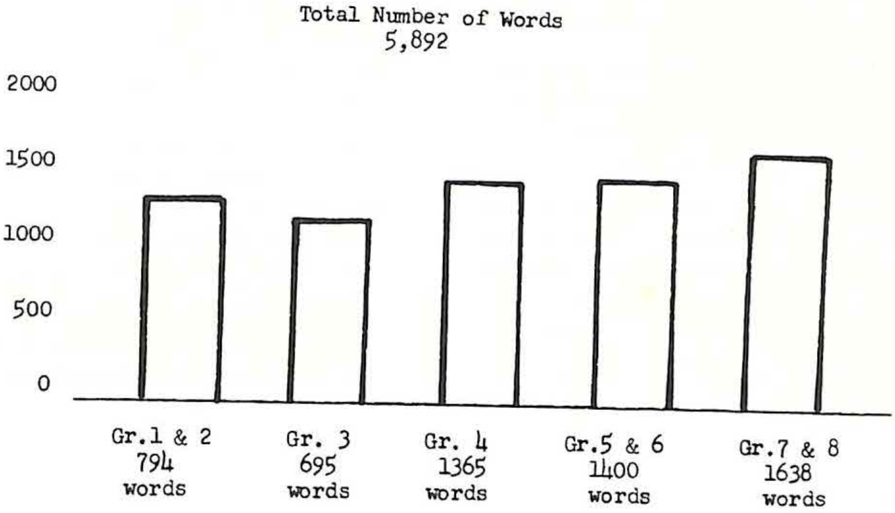
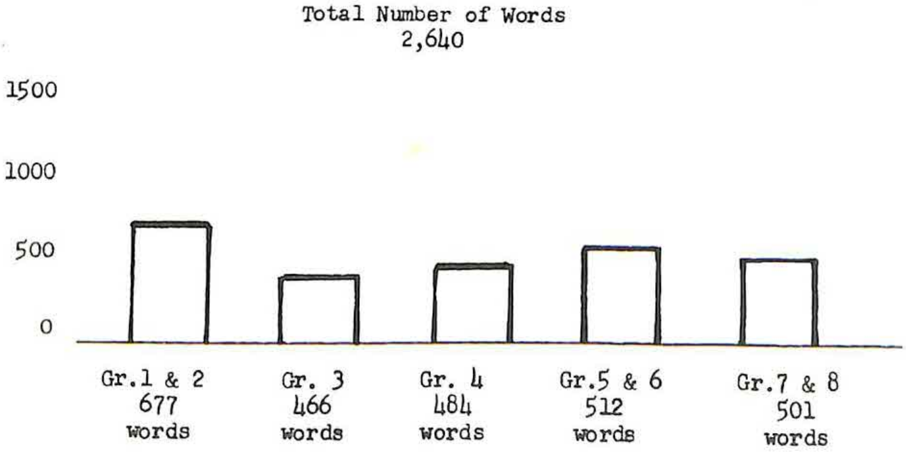


FIGURE 4

NUMBER OF WORDS LISTED FOR EACH GRADE AFTER FIVE OR LESS MEANINGS WERE ELIMINATED



main on the list. This was reasonable because the majority of words have connotations and nuances of meaning that are not concepts and could not be accepted.

The third discarding was:

3. All words having five meanings or less. All of the word lists for each grade were reduced considerably. Little difference exists between the total number of words per grade.

Each word now on the list was looked up in the Thorndike-Lorge *Semantic Count of English Words*. This study appears only in mimeographed form and gives the relative frequency of occurrence of each meaning per mille.

A premise was set up that a meaning to be used in the test had to be at the .050 level. This means that of 1000 occurrences of the word, a meaning was used 50 times. Many meanings were eliminated, but no word was discarded through this screening because a multi-meaning word would have at least one common meaning above the .050 level to be classified in Thorndike's *Basic Vocabulary List*.

The fourth discarding was:

4. Words like do, make, like, come, et cetera, having one over-worked meaning, and the others used infrequently, were discarded. Nearly 50 percent of the words were discarded.

There were still too many words to use in a preliminary test, because each key word selected would test on the average of five meanings, which would mean, if no further discarding was done, that over 7000 meanings would be tested. It would be almost impossible to secure time enough for children in a public school to do such a lengthy project.

The last discarding was:

5. To tabulate the frequency of occurrence of each meaning of the 1572 multi-meaning words. A frequency average percent of meaning was found for each multi-meaning word.

The final list consisted of 321 multi-meaning words and 1984 meanings were measured. Grades 1 and 2 had the highest number of multi-meaning words after they were averaged for highest frequency of occurrence of meaning. Many simple words have high levels of meaning; therefore, the final selection was justifiable.

## Design of Test

Since these tests were developed on the principle of measuring multi-meaning words, the first essential was to find out if pupils recognized words that had more than one meaning.

Sub-Test I was labeled a Multi-Meaning Word Recognition Test. A sample item is shown below:

lettuce moonlight paddle squash port

The pupil underlined the words which have more than one meaning. The preliminary test consisted of four units, 50 items to a unit with five words in each item and a total of 803 multi-meanings were presented to each pupil for recognition. An item analysis was done on this test, and the final tests consisted of three forms, 40 items to a form. A total of 484 multi-meaning words was tested.

These vocabulary tests, except for the recognition test, purport to measure more meanings per word than other vocabulary tests; therefore:

Sub-Test II was a Multi-Meaning Identification Test. A multi-meaning word was listed along with its various common meanings according to highest frequency of occurrence of meaning. In the preliminary test each multi-meaning word was listed with ten alternatives, such as items like these:

| | | |
|---|---|---|
| scale | machine for weighing | (x) |
| | to climb up | (x) |
| | to make a drawing | (x) |
| | according to a graduated table | ( ) |
| | series of tones | (x) |
| | a yardstick | ( ) |
| | a system of numbering | (x) |
| | a barometer | ( ) |

A total of 1984 meanings was measured in this preliminary sub-test.

After the refining was done on each item, the final test was organized into three forms, with 50 items and seven choices for each key word. The total number of meanings tested in the three forms was 690.

A factor of importance in vocabulary testing seems to be the part a word plays in a true contextual reading situation versus a definition knowledge. Multi-meaning words were put in contextual paragraphs. Ninety-six paragraphs were written for the preliminary test, and 294 meanings were tested, for instance:

The stagecoach stopped before the old ( ) lodge. The passengers were glad to see the lights burning as they entered to ( ) lodge for the night. One of the men said that he would ( ) lodge a complaint because he was being delayed on his journey.

1. to find fault 2. airport 3. rather angry
4. to take rest 5. to be weary 6. an inn
7. not to be disturbed

Sixty-one paragraphs were obtained for the

FIGURE 5

NUMBER OF WORDS LISTED FOR EACH GRADE AFTER WORDS HAVING ONLY ONE MEANING OF HIGH FREQUENCY OF OCCURRENCE WERE ELIMINATED

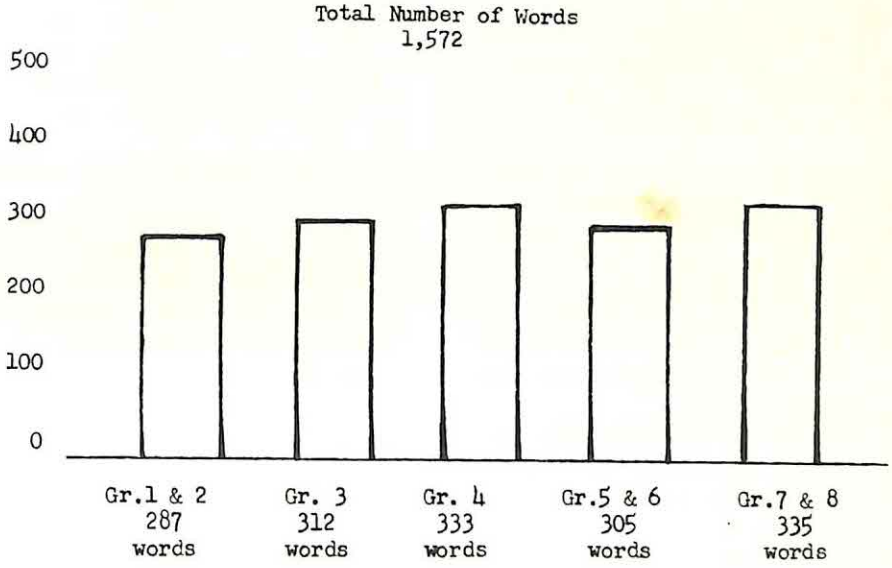
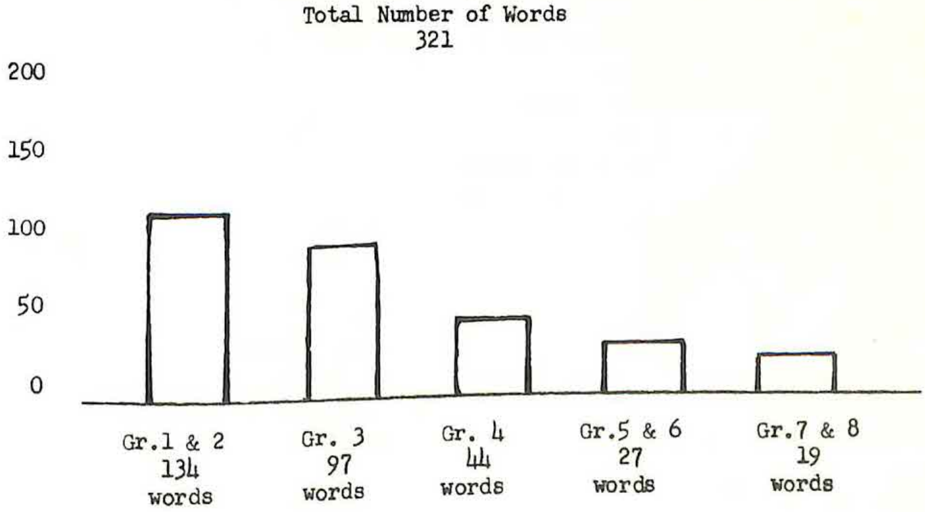


FIGURE 6

NUMBER OF WORDS LISTED FOR EACH GRADE AFTER THE FINAL DISCARDING WAS DONE



three forms of the final test after the item analysis was done, and 183 meanings were measured.

### Trial Administration

Following completion of preliminary editing, trial administrations of the items were conducted. The purpose of these try-outs was two-fold. The first objective was to determine the difficulty of each item for the grades for which the tests were intended. This information would be useful in arriving at a final list of words constituting a test with a difficulty level appropriate for each individual taking these tests. The second purpose was to identify the ineffective items, as such items would have no value and would not be included in the final test.

In order to determine the difficulty of each item and to identify the ineffective items, trial administration of the tests was given to a sample population of 491 pupils in grades 4, 5, 6, 7 and 8 in Dover and Nashua, New Hampshire.

The pages in each test were put in different chronological order so that the first items would not always be first, and the last items would not always be last each time.

The tests were given in twenty-two different sittings of fifteen minutes each for eleven consecutive days, one test in the morning session and one in the afternoon session.

Keys were prepared for each test. All of the tests were hand-scored by persons who had had much experience in correctings.

### Item Analysis

Data showing the distribution of the items according to difficulty level were tabulated.

A frequency count was made on each of the items in the Recognition, Identification, and Context Tests. This was done on the results of the preliminary tests for grades 5, 6 and 7, because it seems that the average would be more consistent in the middle grades for which the test was designed. Items were chosen for the final forms nearest the 50 percent difficulty level and 50 percent ease level, because by this criterion no item would be too easy or too hard for the group tested.

The Walker-Cohen test of significance was used to determine the discriminating power of items. Each word was analyzed, and if the word under consideration had good discriminating powers a much higher percent of pupils with high total scores chose the correct response than the percent of pupils in the lowest total score group choosing the correct response.

The item analysis included:

1. Samples of Recognition Items Accepted and Rejected

Table I represents sample words illustrative of the type of words accepted for the Recognition Test.

A number of words lacked discriminating power, and the average levels of difficulty were either too easy or too hard and were rejected. Samples of these words are listed in Table II.

It was found that from the initial 803 multi-meaning words used in the try-out recognition test, 484 fulfilled the requirements on the difficulty level, as the average centered on the 50 percent difficulty and ease level. The discriminating level of .01 also established that there was only one chance in a hundred the pupil could guess the correct response on these words. They also showed a gradual increase from grade to grade, indicating the higher the grade, the more success the pupils had on the words.

The next item analysis included:

2. Samples of Identification Test Items Accepted and Rejected

To illustrate the item analysis process on 321 key words and 1984 meanings, sample items accepted for the Identification Test are shown in Table III.

Items like the above with gradual levels of difficulty for most of the meanings and good discriminating powers were retained in the final forms. Frequently a meaning fulfilled the requirements except that success at different grade levels varied in range, or there were small minus differences at higher grade levels. It could hardly be expected that responses on so many meanings would always show comparable increases; therefore, meanings were accepted if they showed good discriminating power, and the majority of meanings for the key word had gradual increases of success, and the average difficulty and ease level centered on 50 percent.

Most of the words eliminated in the identification test showed poor discriminating power. Sample items showing this factor are presented in Table IV.

After the item analysis, it was found that 170 words were ineffective, and they were discarded. The number of multi-meaning words used in the final forms of the identification test was 151, and measured 690 meanings.

3. Samples of Paragraph Test Items Accepted or Rejected

An item analysis determined the paragraphs that were desirable from a standpoint of level of ease and difficulty and good discriminating power.

A sample paragraph is presented to show the design of this test, and the number of meanings presented for identification:

> When the village bell rang out the wives of the fishermen rushed to the shore to greet the men. Their faces would ( ) beam as

## TABLE I

### SAMPLE OF RECOGNITION TEST WORDS ACCEPTED ON THE BASIS OF DISCRIMINATING POWER AND GRADUAL EASE OF DIFFICULTY

| | Grades | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | | |
| Word | % | % | % | Average Percent of Difficulty | Discriminating Level |
| Pitch | 37 | 52 | 65 | 51 | .01 |
| Lodge | 45 | 52 | 58 | 52 | .01 |
| Plain | 36 | 41 | 49 | 42 | .01 |
| Cast | 30 | 39 | 59 | 43 | .01 |
| Pick | 33 | 48 | 60 | 47 | .01 |
| Bond | 40 | 47 | 56 | 48 | .01 |
| Trail | 42 | 54 | 68 | 55 | .01 |
| Appeal | 27 | 40 | 65 | 44 | .01 |
| Prime | 29 | 45 | 53 | 42 | .01 |
| Bulb | 31 | 51 | 56 | 47 | .01 |

## TABLE II

### SAMPLES OF RECOGNITION TEST WORDS REJECTED ON THE BASIS OF DISCRIMINATING LEVEL AND DIFFICULTY AND EASE LEVEL

| | Grades | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | | |
| Word | % | % | % | Average Percent of Difficulty | Discriminating Level |
| Latch | 44 | 89 | 82 | 72 | .23 |
| Lime | 28 | 36 | 35 | 33 | .11 |
| Quail | 17 | 27 | 32 | 25 | .11 |
| Wax | 16 | 41 | 37 | 37 | .23 |
| Harp | 33 | 56 | 70 | 53 | .39 |
| Elder | 11 | 41 | 42 | 31 | .23 |
| Ham | 58 | 78 | 82 | 73 | .39 |
| Organ | 12 | 26 | 48 | 29 | .11 |
| Yoke | 15 | 37 | 32 | 28 | .23 |
| Barrier | 11 | 21 | 42 | 25 | .11 |

## TABLE III

### SAMPLES OF IDENTIFICATION TEST ITEMS ACCEPTED ON THE BASIS OF DISCRIMINATING POWER AND GRADUAL EASE OF DIFFICULTY

| | Grades | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | | |
| Key Word (Palm) | % | % | % | Average Percent of Difficulty | Discriminating Level |
| Meanings Tested | | | | | |
| Tropical tree | 61 | 70 | 72 | 68 | .01 |
| Emblem of victory | 24 | 33 | 44 | 34 | .01 |
| Flat part of hand | 57 | 68 | 81 | 69 | .01 |
| Measure of length | | | | | ... |
| To pass by fraud | 16 | 29 | 34 | 26 | .01 |
| | Average difficulty of all meanings | | | 49 | |

| | Grades | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | | |
| Key Word (Crown) | % | % | % | Average Percent of Difficulty | Discriminating Level |
| Meanings Tested | | | | | |
| A royal headdress | 82 | 86 | 94 | 87 | .01 |
| Topmost part of head | 51 | 55 | 57 | 54 | .01 |
| British coin worth five shillings | 11 | 18 | 16* | 12 | .01 |
| Part of tooth | 25 | 27 | 24 | 25 | .39 |
| To invest with regal power | 19 | 24 | 26 | 23 | .01 |
| To hit on head | 42 | 36 | 52 | 43 | .11** |
| Crest of a bird | 29 | 27 | 46 | 34 | .03** |
| To curve upward | 13 | 15 | 22 | 17 | .11** |
| | Average difficulty of all meanings accepted | | | 45 | |

---

* Very few key words had every meaning showing gradual gain, but discriminating power was acceptable so it was included in the final test.

**Meaning excluded on the basis that it lacked discriminating power.

## TABLE IV

## SAMPLE OF IDENTIFICATION TEST ITEMS REJECTED ON THE BASIS OF POOR DISCRIMINATING POWER SHOWING NO TRUE DIFFERENCES, AND DIFFICULTY LEVEL NOT NEAR ENOUGH TO 50 PERCENT

| | Grades | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | | |
| Key Word (Address) | % | % | % | Average Percent of Difficulty | Discriminating Level |
| Meanings Tested | | | | | |
| To make a speech | 15 | 41 | 57 | 38 | .01 |
| To woo | 21 | 41 | 44 | 35 | .11 |
| A person's place of | 90 | 91 | 92 | 91 | .04 |
| Manners and bearing | 18 | 27 | 31 | 25 | .04 |
| To apply or devote oneself to a task | 17 | 17 | 22 | 19 | .23 |
| To take aim at a ball | 13 | 28 | 26 | 22 | .23 |
| | Average difficulty of meanings | | | 38 | |
| Key Word (Common) | | | | | |
| Meanings Tested | | | | | |
| Shared by more than one | 26 | 41 | 46 | 36 | .01 |
| Usual | 10 | 19 | 22 | 17 | .01 |
| Low, vulgar | 17 | 25 | 29 | 24 | .11 |
| Tract of public land | 28 | 52 | 54 | 45 | .01 |
| Average in rank | 11 | 18 | 22 | 17 | .01 |
| | Average difficulty of meanings | | | 28 | |

## TABLE V

### SAMPLE OF CONTEXT TEST ITEMS ACCEPTED ON THE BASIS OF DISCRIMINATING POWER AND GRADUAL EASE OF DIFFICULTY

| Multi-Meaning Word Used in Paragraph (beam) | Grades 5 % | Grades 6 % | Grades 7 % | Average Percent of Difficulty | Discriminating Level |
|---|---|---|---|---|---|
| Meanings Tested | | | | | |
| A heavy timber | 29 | 38 | 55 | 41 | .01 |
| Smile radiantly | 51 | 72 | 76 | 66 | .01 |
| A ray of light | 49 | 58 | 72 | 60 | .01 |
| | Average difficulty of accepted meanings | | | 55 | |

## TABLE VI

### SAMPLE OF CONTEXT TEST PARAGRAPH REJECTED ON THE BASIS OF POOR DISCRIMINATING POWER

| Multi-Meaning Word Used in Paragraph (bar) | Grades 5 % | Grades 6 % | Grades 7 % | Average Percent of Difficulty | Discriminating Level |
|---|---|---|---|---|---|
| Meanings Tested | | | | | |
| To prevent | 27 | 46 | 52 | 42 | .11 |
| Barrier | 26 | 45 | 63 | 45 | .23 |
| Bank of Sand | 41 | 60 | 71 | 57 | .09 |
| | Average difficulty of item | | | 48 | |

they saw the sails in the distance.. A ( ) beam of light would shine in the water, and later they could make out the nets hanging from a ( ) beam on the stern.

1. a radio 2. a heavy timber 3. show worry 4. smile radiantly 5. candlelight 6. heavy cable 7. a ray of light

The item analysis determined the paragraphs that were desirable from a standpoint of level of ease and difficulty and good discriminating power. Table V includes responses on the above paragraph, and were the type accepted for the final forms of the test.

An illustration is given in Table VI of a paragraph that was rejected on the basis of poor discriminating power.

As the flood rushed toward the old dam the men worked desperately to ( ) bar it from overflowing onto the highway. Sandbags were used as a ( ) bar where the dam was weak. Completely out of sight was the ( ) bar that usually reached out into the channel.

1. a rushing sound 2. sand bank 3. prevent 4. danger signs 5. support 6. make bare 7. barrier

Paragraphs like the above had to be discarded because the responses were influenced too strongly by the chance factor.

The number of paragraphs in the preliminary context test was ninety-six. Sixty-one paragraphs fulfilled the requirements set up for items to be used in the final context tests.

## Test Developments

After each item was analyzed, the final forms were made, using all the words that fulfilled the requirements postulated for good test items.

In order to ensure balanced items in each form of the three tests, the words were grouped according to average levels of difficulty and ease. The items were arranged in blocks of five, and an average was taken after every fifth item to equalize each form. Thus, pupils who would not finish the test would always encounter similar items in any form of the tests at every fifth point.

The words were placed as nearly as possible in each form to measure the same number of meanings. Occasionally this was impossible because if they balanced in difficulty, they did not in number of meanings at every fifth item. When this occurred, they were balanced in the next block of five items.

Table VII shows in schematic form the plan for assigning words to each form.

## Statistical Data on Population

To secure data bearing upon the type of population reliability and validity of the constructed tests, the following tests were given to 491 pupils in grades 4 to 8 in the public schools at Dover and Nashua, between November 1 and December 19, 1951.

1. California Short-Form Test of Mental Maturity, Elementary S-Form, Grades 4-8 (1938).
2. Gates Reading Survey for Grades 3 (2nd Half) to 10 (Vocabulary, Level of Comprehension, Speed and Accuracy).
3. Multi-Meaning Vocabulary Tests (Recognition, Identification, and Context), Form A.
4. Multi-Meaning Vocabulary Tests (Recognition, Identification, and Context), Form B.
5. Multi-Meaning Vocabulary Tests (Recognition, Identification, and Context), Form C.

## Analysis of Data

The sensitivity of any test as a measuring instrument depends upon the ascending order of means between grades. Every test should serve to distinguish relative positions of groups as regards the trait which is measured. Table VIII illustrates the differences between the mean raw scores and standard deviations of each grade taking the recognition test.

The greatest apparent differences in mean scores were between grade 4 and grade 5, an average difference of 21 points; then grade 5 and grade 6, an average difference of 10 points, and grade 6 and grade 7, an average difference of seven points. Grade 8's difference was three points lower on Form A, and practically equal on the other two forms to grade 7.

Data showing the mean scores for grade 4 through grade 8 obtained on the identification test are shown in Table IX.

Whereas, the mean scores for each grade show an increase from one grade to the next, an average difficulty of 5 points between grade 4 and grade 5, of 21 points between grade 5 and grade 6, of 13 between grade 6 and grade 7, the average difficulty between grade 7 and grade 8 drops to 4.

Data showing the mean scores obtained on the context by pupils in grade 4 through grade 8 are represented in Table X.

A striking fact lies in the failure of grade 8 to make little or no gains on any of the tests.

The gains for the other grades were consistent and a steady increase was noted between mean scores of consecutive grade levels.

The final forms of the tests were not administered, but re-scoring was done on each of the items accepted for the final tests from the try-out tests. The advantages of re-scoring after tests have been refined are:

## TABLE VII

## EQUALITY OF ITEMS FOR THE THREE FORMS OF THE TEST

| Word | Average Percent of Difficulty | Discriminating Level | Number of Meanings |
|---|---|---|---|
| | | Form A | |
| 1. Palm | 48 | .01 | 4 |
| 2. Beat | 40 | .01 | 5 |
| 3. Snap | 49 | .01 | 5 |
| 4. Mill | 56 | .01 | 4 |
| 5. Fly | 50 | .01 | 5 |
| | | Form B | |
| 1. Bolt | 47 | .01 | 5 |
| 2. Master | 42 | .01 | 4 |
| 3. Point | 47 | .01 | 5 |
| 4. Arm | 58 | .01 | 4 |
| 5. Net | 49 | .01 | 4 |
| | | Form C | |
| 1. Date | 50 | .01 | 4 |
| 2. Guard | 48 | .01 | 5 |
| 3. Order | 43 | .01 | 5 |
| 4. Match | 52 | .01 | 4 |
| 5. Grant | 50 | .01 | 4 |

Totals:

Form A - Average Percent of Difficulty, 49; Discriminating Level, .01; Number of Meanings, 23.

Form B - Average Percent of Difficulty, 49; Discriminating Level, .01; Number of Meanings, 2.

Form C - Average Percent of Difficulty, 49; Discriminating Level, .01; Number of Meanings, 22.

Note: Inspection of the table shows that the items were balanced in dificulty and discriminating levels.

## TABLE VIII

### COMPARISON OF MEANS, STANDARD DEVIATIONS ON THE THREE FORMS OF THE RECOGNITION TEST FOR GRADES 4 - 8

| | Form A | | Form B | | Form C | |
|---|---|---|---|---|---|---|
| | Mean | S. D. | Mean | S. D. | Mean | S. D. |
| Grade 4 | 61.201 | 28.77 | 64.294 | 29.74 | 66.458 | 28.96 |
| Grade 5 | 83.186 | 32.47 | 86.111 | 30.96 | 86.217 | 33.17 |
| Grade 6 | 96.520 | 31.82 | 94.570 | 30.12 | 96.419 | 29.70 |
| Grade 7 | 102.255 | 23.71 | 102.153 | 22.13 | 102.969 | 20.92 |
| Grade 8 | 99.758 | 31.14 | 102.542 | 29.11 | 102.232 | 30.51 |

## TABLE IX

### COMPARISON OF MEANS, STANDARD DEVIATIONS ON THE THREE FORMS OF THE IDENTIFICATION TEST FOR GRADES 4 - 8

| | Form A | | Form B | | Form C | |
|---|---|---|---|---|---|---|
| | Mean | S. D. | Mean | S. D. | Mean | S. D. |
| Grade 4 | 72.372 | 30.10 | 74.510 | 30.03 | 75.634 | 30.38 |
| Grade 5 | 77.934 | 35.70 | 79.863 | 33.45 | 81.773 | 34.64 |
| Grade 6 | 98.035 | 30.11 | 101.773 | 31.19 | 99.550 | 30.06 |
| Grade 7 | 110.725 | 25.46 | 113.480 | 29.77 | 113.582 | 27.19 |
| Grade 8 | 115.118 | 35.34 | 117.283 | 34.64 | 117.840 | 37.12 |

### TABLE X

### COMPARISON OF MEANS, STANDARD DEVIATIONS ON THE THREE FORMS OF THE CONTEXT TEST FOR GRADES 4 - 8

| | Form A | | Form B | | Form C | |
|---|---|---|---|---|---|---|
| | Mean | S. D. | Mean | S. D. | Mean | S. D. |
| Grade 4 | 20.659 | 9.15 | 21.341 | 10.22 | 22.876 | 10.21 |
| Grade 5 | 26.242 | 12.21 | 27.005 | 12.22 | 27.858 | 12.65 |
| Grade 6 | 31.748 | 12.97 | 31.240 | 13.11 | 31.192 | 14.11 |
| Grade 7 | 35.469 | 11.51 | 35.724 | 10.97 | 35.113 | 11.86 |
| Grade 8 | 35.711 | 12.54 | 34.890 | 12.80 | 33.701 | 13.14 |

### TABLE XI

### CORRELATIONS* BETWEEN SCORES ON THE CONSTRUCTED MULTI-MEANING SUB-TESTS AND MENTAL AGE OF GRADES 4 - 6 - 8 PUPILS

| Factor | Grade | No. Cases | Recognition | Identification | Context |
|---|---|---|---|---|---|
| M. A. | 4 | 97 | .644 ± .06 | .703 ± .05 | .649 ± .06 |
| M. A. | 6 | 99 | .748 ± .04 | .787 ± .04 | .810 ± .04 |
| M. A. | 8 | 98 | .777 ± .04 | .812 ± .03 | .821 ± .03 |

*Pearson product-moment

## TABLE XII

## CORRELATIONS BETWEEN SCORES ON THE CONSTRUCTED MULTI-MEANING SUB-TESTS AND READING AGE FOR GRADES 4 - 6 - 8 PUPILS

| Factor | Grade | No. Cases | Recognition | Identification | Context |
|---|---|---|---|---|---|
| R. A. | 4 | 97 | .778 ± .04 | .816 ± .03 | .830 ± .03 |
| R. A. | 6 | 99 | .873 ± .02 | .887 ± .02 | .876 ± .02 |
| R. A. | 8 | 98 | .826 ± .03 | .881 ± .02 | .871 ± .02 |

## TABLE XIII

## CORRELATIONS BETWEEN SCORES ON THE CONSTRUCTED MULTI-MEANING SUB-TESTS AND CHRONOLOGICAL AGE OF GRADES 4 - 6 - 8 PUPILS

| Factor | Grade | No. Cases | Recognition | Identification | Context |
|---|---|---|---|---|---|
| C. A. | 4 | 97 | -.257 ± .03 | -.255 ± .05 | -.277 ± .02 |
| C. A. | 6 | 99 | -.117 ± .10 | -.453 ± .08 | -.430 ± .08 |
| C. A. | 8 | 98 | -.115 ± .10 | -.343 ± .09 | -.441 ± .08 |

### TABLE XIV

### CORRELATIONS BETWEEN SCORES ON THE CONSTRUCTED MULTI-MEANING SUB-TESTS AND GATES VOCABULARY TEST FOR GRADES 4 - 6 - 8

| Test | Grade | No. Cases | Recognition | Identification | Context |
|---|---|---|---|---|---|
| Gates Vocabulary | 4 | 97 | .755 ± .04 | .804 ± .04 | .799 ± .04 |
| Gates Vocabulary | 6 | 99 | .876 ± .02 | .883 ± .02 | .881 ± .02 |
| Gates Vocabulary | 8 | 98 | .826 ± .03 | .882 ± .02 | .871 ± .03 |

### TABLE XV

### CORRELATIONS BETWEEN SCORES ON THE CONSTRUCTED MULTI-MEANING SUB-TESTS AND THE GATES COMPREHENSION TEST FOR GRADES 4 - 6 - 8

| Test | Grade | No. Cases | Recognition | Identification | Context |
|---|---|---|---|---|---|
| Gates Compre. | 4 | 97 | .716 ± .05 | .754 ± .05 | .758 ± .03 |
| Gates Compre. | 6 | 99 | .759 ± .04 | .801 ± .04 | .791 ± .04 |
| Gates Compre. | 8 | 98 | .802 ± .04 | .811 ± .04 | .871 ± .03 |

### TABLE XVI

### CORRELATIONS BETWEEN THE THREE SUB-TESTS OF THE MULTI-MEANING VOCABULARY TEST FOR GRADES 4 - 6 - 8

| Grade | No. Cases | Tests I-II | Tests I-III | Tests II-III |
|---|---|---|---|---|
| 4 | 97 | .828 ± .03 | .834 ± .03 | .821 ± .03 |
| 6 | 99 | .885 ± .02 | .897 ± .02 | .885 ± .02 |
| 8 | 98 | .878 ± .02 | .817 ± .03 | .861 ± .03 |

### TABLE XVII

### CORRELATIONS BETWEEN SCORES ON THE THREE FORMS OF THE MULTI-MEANING VOCABULARY TEST FOR GRADES 4 - 6 - 8

| Grade | No. Cases | Forms A - B | Forms A - C | Forms B - C |
|---|---|---|---|---|
| 4 | 97 | .850 ± .03 | .812 ± .04 | .813 ± .04 |
| 6 | 99 | .831 ± .03 | .898 ± .02 | .822 ± .03 |
| 8 | 98 | .872 ± .03 | .827 ± .03 | .869 ± .03 |

1. The results were obtained from the same population without having them take the test a second time.
2. No complete form was given at the same time as the items had not been selected for the final tests when the preliminary tests were given.
3. All the tests were corrected a second time.

### Validity of the Tests

In determining the validity of any test, it is necessary to find the correlations between the criteria used.

California (p. 110)
Reading Age
Chronological Age
Gates Reading Survey for Grade 2 (2nd half to 10)
(Vocabulary, Level of Comprehension, Speed and Accuracy)

Table XI gives the coefficient of correlation found between mental age secured from the above mental test and the scores on the sub-tests of the Multi-Meaning Vocabulary Tests.

It is recognized that the criterion of mental age represents as satisfactory an independent index of the ability necessary for vocabulary as can be secured.

The use of another factor was made to assure the validity of the multi-meaning tests. Reading ages secured from the results of the Gates Reading Survey were obtained. Table XII shows the correlation found between this factor and the constructed tests.

The correlation with reading age is high, since the task of identifying and knowing the definition of words is a component of reading. Table XIII presents coefficients of correlations between chronological age and the scores obtained on the sub-tests of the Multi-Meaning Vocabulary Tests.

The correlations between chronological age and the multi-meaning vocabulary sub-tests in each instance were found to be negative. This finding denotes many of the oldest pupils in the sample obtained the lowest scores and many of the youngest pupils had the highest scores.

To secure additional data bearing upon the validity of the constructed tests, correlations between subtests in the Gates Reading Survey Test were found. This test is divided into three subtests and a correlation between the vocabulary test and the multi-meaning vocabulary tests was made.

Table XIV presents data on the correlations that have been found.

All correlations between the Gates Vocabulary Test and the multi-meaning sub-tests are high. This finding indicates that the pupils who made high scores in the Gates Vocabulary Test made high scores on the constructed tests while pupils making low scores on the Gates Vocabulary Test made low scores on the multi-meaning Vocabulary Tests.

With the total scores obtained on the Gates Comprehension Test, a correlation was made with the Multi-Meaning Sub-tests. Table XV shows the findings.

Gates Comprehension Test and the Multi-Meaning Vocabulary Sub-Tests are all statistically high. It is significant that these two Gates sub-tests correlate about equally well with the constructed multi-meaning vocabulary tests, and that there are consistently high correlations.

### Reliability of the Tests

Basic statistical procedures were necessary to determine the reliability of the constructed tests. Coefficients of reliability for the sub-tests were found. Table XVI shows the relationships on these three sub-tests.

It was to be expected that pupils who obtained a high score on one of the sub-tests would be likely to secure a high score on a similar test which measured a like skill differing only in application of technique.

The large number of multi-meaning words used in the preliminary tests made it possible after the test was refined to organize the material into three forms. Each form included a recognition, identification, and context test. In order to ascertain whether or not the pupils obtained a corresponding score on each form of the test, coefficients of reliability were found. Table XVII presents a comparison of the correlations obtained on Forms A, B, and C of the Multi-Meaning Vocabulary Test.

High correlations were found among all the forms of the Multi-Meaning Vocabulary Test. In view of the fact that each form was equally weighted in difficulty, discriminating level, and number of meanings, it is not surprising that such high coefficients of reliability were found.

### Conclusions

Basic statistics for the evaluation of this test were presented. The following facts are apparent:

1. More words have been tested in approximately the same time and space as a single meaning vocabulary test.
2. The tests are highly valid.
3. The different forms and the sub-tests show very high reliability.
4. A more substantial inventory of pupils' knowledge of words which have more than one meaning is secured.

## REFERENCES

California Short-Form of Mental Maturity, Elementary S-Form (Los Angeles: California Test Bureau, 1950).

Gates Reading Survey for Grades III to X (New York: Bureau of Publications, Teachers College, Columbia University, 1942).

Metropolitan Achievement Tests, Elementary Battery, Form R (Yonkers: World Book Co., 1946).

Thorndike, Edward L., and Irving Lorge. The Teacher's Word Book of 30,000 Words (New York: Bureau of Publications, Teachers College, Columbia University, 1955).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 2, December 1959)

# THE DEVELOPMENT OF THE EARLY SCHOOL PERSONALITY QUESTIONNAIRE*

RICHARD W. COAN
University of Arizona
RAYMOND B. CATTELL
University of Illinois

IN THE course of an extensive investigation of personality in middle childhood, a 200-item group questionnaire was constructed. It was administered to two samples of first- and second-grade children. One sample consisted of 151 children from the public schools of Mahomet and Rantoul, Illinois. The other consisted of 181 children in Decatur, Illinois. The item data from the two samples were subjected independently to factor analysis. Each study culminated in 18 obliquely rotated factors. This work was reported in two earlier articles (4, 7). In one of these articles (7), the alignment of the two sets of factors was considered. When a combination of criteria was applied, it was found that there was reasonably clear matching for 11 factors from either study with a corresponding set of 11 factors from the other study.

## Construction of the ESPQ Scales

It was decided that the eleven matching factor pairs represented the most stable, reproducible factors underlying the questionnaire data, and these were accordingly chosen for inclusion in a new factored questionnaire test. In addition, a large-variance factor appearing in the later, but not the earlier, study was accepted for representation in this test. This decision seemed justified by the fact that the later study utilized a larger sample and yielded a closer approximation to simple structure. In addition, it was thought desirable to include a thirteenth factor of general intelligence. Because of the nature of the original item battery, of course, such a factor had not appeared in either analysis.

For each of the 12 personality factors, two six-item scales were constructed. Each scale met the usual criteria for factor-scale construction: a) adequate representation of the given factor in terms of item correlations with the given reference vector in both studies, b) suppression, or balancing out, of irrelevant factor variance, c) an equality of a and b keyed alternatives.

For intelligence, two eight-item scales were constructed. To insure adequate breadth, each scale was assigned items of four different types: synonyms, reasoning, subordination, and superordination. These were designed to blend in form with the personality items, and they utilize the same two-alternative response form. In difficulty level, they were aimed at the six-to-eight-year level. In the course of standardization, data will also be gathered on a battery of alternative items so that any necessary item replacements can be made.

Thirteen pairs of scales were thus constructed. These were segregated into two questionnaire forms (Forms A and B) for a test which has been designated the Early School Personality Questionnaire (ESPQ). Each form contains 12 six-item personality scales and one eight-item intelligence scale, making a total of 80 items. The correspondence of the factor scales, considered in final numerical order, to the factors in the two studies is shown in the following list (where "D" refers to the study of Decatur children and "MR" to the earlier study based on data from Mahomet and Rantoul):

1. D1, MR 2 (-)
2. D2, MR 4
3. D3, MR 3
4. D4, MR 1
5. D5
6. D6, MR 5
7. D8, MR 6
8. D7, MR 15
9. D9, MR 10 (-)
10. D10, MR 11 (-)
11. D13, MR 7 (-)
12. D15, MR 14
13. Intelligence

Because of the brevity of the scales, it is assumed that neither form will be used in isolation, except in rare circumstances. The division into two forms, however, should facilitate two-session administration. Each form will require 30 or 40

---

*The research reported in this article was supported by a grant from the Department of Public Welfare of the State of Illinois.

minutes of testing time. Administration of the two forms at a single session would be advisable only with older subjects who can accept a fairly rapid pace.

An answer form and a key have been specially designed to permit easy and rapid scoring. For each form, there is a four-page answer form. Each page contains twenty rectangular boxes arranged in two columns. In the middle of each box is an item number, as well as a picture which can serve the same purpose for younger subjects. There is an A at the left end of each box and a B at the right end.

The standard instructions are an adaptation of those employed in the original researches. For each item, the examiner first directs the class to the appropriate box and then reads the item. Marking instructions are given to identify each alternative with the A or the B. The children are asked to indicate their responses by drawing a line through either the A or the B in each box.

To give the reader some idea of the content of each scale, we shall list the questions utilized for each factor scale. The keyed alternative corresponding to the high end of the scale is underlined.

### Factor 1

#### Form A

1. Would you rather play: a) school, or b) cowboys and Indians?
2. Which would you rather be: a) a doctor, or b) a teacher?
11. Would you rather: a) fly an airplane, or b) be a teacher?
21. Do you like to talk to teachers? a) yes, or b) no.
31. Do you ever talk back to your mother? a) yes, or b) no.
32. Would you rather: a) go to a party, or b) stay home and play?

#### Form B

1. If you were in a play, would you rather be: a) a teacher, or b) a hunter?
2. If another child has your coat, do you: a) take it away from him, or b) tell the teacher?
11. a) Do you like to see other children cry, or b) does it make you sad?
21. Would you rather: a) be in a play, or b) make something out of wood?
31. Do you ever feel like running away from home? a) yes, or b) no.
32. Would you rather have: a) a new baby come to live with you, or b) a little dog come to live with you?

### Factor 2

#### Form A

3. Do you shiver when you hear a squeaky door or chalk scraping on the blackboard? a) Yes, or b) no.
12. Do you like: a) all the things your mother gives you to eat, or b) only some of them?
13. When your mother tells you you can't do something, do you want to do it even more? a) Yes, or b) no.
22. Does it bother you if people say you don't listen? a) Yes, or b) no.
23. When you are told to do something or put something away: a) do you always do it right away, or b) do you sometimes forget what you are supposed to do?
33. a) Do you have a lot of fun, or b) do things sometimes go wrong?

#### Form B

3. Do grown-ups ever say you daydream too much? a) Yes, or b) no.
12. Would you rather: a) play by yourself, or b) play with other boys and girls?
13. Do you ever feel like crying when you see something sad in the movies? a) Yes, or b) no.
22. Do people ever call you naughty and mischievous? a) Yes, or b) no.
23. a) Are other children always nice to you, or b) do they sometimes pick on you?
33. a) Are you happy all the time, or b) do you sometimes get sad?

### Factor 3

#### Form A

4. a) Are your dreams usually nice, or b) do they scare you?
5. Would you rather: a) go on a long trip in the car, or b) go to school?
14. When your friends want to play a different game, do you: a) play their game, or b) go on with your game?
24. When you lose a book, do you: a) cry, or b) just laugh?
34. Which do you like better: a) funny books, or b) the books you have in school?
35. Would you rather tell your mother and father: a) about school, or b) about a game you played with your friends?

#### Form B

4. Do you like to play: a) hard games that you can win, or b) easy games where nobody wins?

5. Would you rather: a) work at home, or b) go to school?
14. Do people ever say you talk too much or call you a chatterbox? a) Yes, or b) no.
24. Do you think: a) school is hard, or b) school is easy?
34. Would you rather: a) watch a game, or b) learn something in school?
35. a) Do you think grown-ups should listen to you and help you more than they do, or b) do you like it better when they just leave you alone?

## Factor 4

### Form A

6. Would you rather have: a) a friend who can read well, or b) a friend who is good at ball games?
15. Would you rather: a) climb a tree, or b) look at a book?
16. Do you think: a) everybody likes you, or b) only some people?
25. Would you rather look at: a) comic strips that are funny, or b) comic strips with a lot of fighting and shooting in them?
26. Do people ever say you're stuck-up? a) Yes, or b) no.
36. a) Can you touch a big bug, or b) are you afraid to touch bugs?

### Form B

6. Would you rather: a) color a book, or b) climb a tree?
15. Would you rather: a) hunt for birds, or b) draw pictures of birds?
16. Would you rather: a) watch people dancing, or b) hear a story about airplanes?
25. Would you rather: a) talk to a friend, or b) look at funny books?
26. Which do you like better: a) home, or b) school?
36. Do you like to climb trees? a) Yes, or b) no.

## Factor 5

### Form A

7. Would you rather: a) talk to your friends, or b) talk to your teachers?
8. Would you rather: a) stay at home, or b) go shopping with your mother?
17. a) Do you like to talk to your teacher, or b) are you sometimes a little afraid to?
27. When your friends start fighting: a) do you just leave them alone, or b) do you tell them to stop it?
37. When you get angry: a) do you shout, or b) just cry?
38. Are you as good-looking as the other children in your class? a) Yes, or b) no.

### Form B

7. Would you rather: a) play a noisy game where you pretend to be wild animals, or b) listen to a story read by your teacher?
8. When somebody bawls you out: a) do you cry, or b) do you get mad?
17. When your mother and father are busy, do you like to help them? a) Yes, or b) no.
27. If you had to make your bed: a) would you listen to the radio first and then make it, or b) would you make it right away?
37. a) Do you like to talk to your father, or b) would you rather not talk to him?
38. Do you like to cross busy streets? a) Yes, or b) no.

## Factor 6

### Form A

9. a) Do you like for other children to play with your toys, or b) does it sometimes bother you?
18. Do you cry: a) more than other boys and girls, or b) less?
19. Do people boss you around too much? a) Yes, or b) no.
28. a) Are you always very careful how you move, or b) do you sometimes rush around when you play and knock things over?
29. Would you rather be: a) a grown-up, or b) a baby?
39. Do your friends sometimes say: a) bad things about you, or b) only true things?

### Form B

9. If you get upset or sad: a) do you get happy again pretty soon, or b) do you stay sad for a long time?
18. When you get hurt, do you: a) cry, or b) try to keep from crying?
19. Do people sometimes punish you when you haven't done anything wrong? a) Yes, or b) no.
28. Do you like to play: a) new games, or b) just games you already know?
29. When you get hurt: a) do you just try to forget about it, or b) do you sometimes cry?
39. Is it sometimes hard to get people to understand what you're saying? a) Yes, or b) no.

## Factor 7

### Form A

41. Would you rather: a) listen to a story, or b) watch two dogs fight?

42. Who usually has better ideas: a) you, or b) your friends?
51. Are you pretty good at: a) everything, or b) just a few things?
61. a) Are you getting along pretty well, or b) do you have a lot of problems and troubles?
71. Do you like to tell stories to other people? a) Yes, or b) no.
72. Do you usually: a) do what other people say, or b) do what you want to do?

Form B

41. Which do you like better: a) hearing stories about what a boy or girl does, or b) really doing things yourself?
42. a) Does your mother let you do almost anything you want, or b) are there lots of things she won't let you do?
51. a) Are you stronger than other children, or b) are they stronger than you?
61. When you argue with people: a) do you sometimes find out that you were wrong, or b) are you nearly always right?
71. When you want to say something: a) do you just say it, or b) do you think it over first?
72. a) Do you have to do things you don't want to do, or b) do you always do what you want to?

### Factor 8

Form A

43. Do you like: a) to play with dogs, or b) to stay away from them?
52. Would you rather: a) go to a party, or b) listen to the radio and TV?
53. Do you get into: a) more trouble than most other children, or b) less trouble?
62. a) Do you like to do things the way your mother says, or b) is your own way better?
63. When children play tricks on you: a) do you cry, or b) do you get mad?
73. Which do you like better: a) cats, or b) dogs?

Form B

43. Which would you rather be: a) a dog, or b) a cat?
52. a) Are you always neat and tidy, or b) are you sometimes careless and messy?
53. Which would you like better: a) a story about fighting Indians, or b) a story about how Indians made clothing?
62. Would you rather: a) run, or b) sit still?
63. Would you rather do something for: a) your mother, or b) your father?
73. a) Do you have days when everything goes wrong, or b) are you happy all the time?

### Factor 9

Form A

44. Does anyone ever call you a cry-baby? a) Yes, or b) no.
45. a) Are you always pretty lucky, or b) do more bad things happen to you than to other children?
54. Is your teacher: a) nicer to other children, or b) just as nice to you?
64. On the playground: a) do you make a lot of noise, or b) are you mostly quiet?
74. Would you rather: a) build things with your friends, or b) build things by yourself?
75. Do you think people ever say bad things about you behind your back? a) Yes, or b) no.

Form B

44. Do you know any children who are so dumb that it's no fun to play with them? a) Yes, or b) no.
45. Do you like to: a) run, or b) just walk?
54. On the playground: a) do you play by yourself, or b) mostly with other children?
64. a) Do you have a lot of friends, or b) just a few friends?
74. On the playground: a) do you run most of the time, or b) stand still a lot?
75. Do you like to: a) help your mother and father with things, or b) just play?

### Factor 10

Form A

46. Would you rather go on a trip with: a) your father, or b) your mother?
55. If you wake up in the dark: a) do you sometimes feel scared, or b) do you like it because it's so dark and quiet?
56. Would you rather: a) play a noisy game, or b) look at a book by yourself?
65. When people are talking about a movie that you have seen: a) do you want to tell it your way, or b) just listen to them?
66. When your mother is angry: a) do you feel like crying, or b) do you feel happy anyway?
76. If you were up on a big rock: a) would you be scared, or b) would you just laugh?

Form B

46. Would you rather play with: a) older children, or b) younger children?
55. Which would you like better: a) to hear stories about bears, or b) to have bears here right now?
56. Would you rather talk to: a) your father, or b)

your mother?

65. When you go to bed: a) do you lie awake for a long time, or b) do you go to sleep right away?
66. When you start to say something: a) do grown-ups always listen to you, or b) do they do all the talking?
76. Do you sometimes feel a little scared when you're up on a high place? a) Yes, or b) no.

### Factor 11

Form A

47. When you get a new toy: a) do you let other children play with it, or b) do you watch out that no one breaks it?
48. Does your mother ever go away for a little while and leave you at home all by yourself? a) Yes, or b) no.
57. Do you like: a) a friend who talks a lot, or b) one who is quiet?
67. Do you ever take your toys to bed with you? a) Yes, or b) no.
77. Would you rather play: a) in your own yard, or b) in somebody else's?
78. Are you ticklish? a) Yes, or b) no.

Form B

47. Can you do things: a) better than most boys and girls, or b) not as well as most boys and girls?
48. a) Do people sometimes bawl you out when you haven't done anything wrong, or b) do they always treat you the way they should?
57. a) Does someone wake you up in the morning, or b) do you wake up by yourself?
67. Do you ever talk in your sleep and wake yourself up? a) Yes, or b) no.
77. a) Do you wish you had more friends to play with, or b) do you have enough friends?
78. When your mother and father say it's time for bed: a) do you like to go to bed, or b) do you want to stay up longer?

### Factor 12

Form A

49. Do you like: a) to meet new boys and girls, or b) to be with children you already know?
58. Do you like: a) to tell other children what to do, or b) do what other children want to do?
59. Do you ever do things that you should not do? a) Yes, or b) no.
68. Do you like to play games: a) with a lot of children, or b) just one or two children that you know?
69. a) Can you remember stories, or b) do you forget them very soon?
79. Would you rather: a) make something with blocks or bricks, or b) play with other children?

Form B

49. a) Do you make your bed in the morning, or b) does your mother make your bed?
58. When you get angry: a) do you sometimes yell and stamp your feet, or b) do you just try to forget about it?
59. Would you rather: a) look at a picture book by yourself, or b) look at it with another boy or girl?
68. When children fight, do you think grown-ups sometimes punish the wrong one? a) Yes, or b) no.
69. Do you: a) usually put your clothes away at night, or b) just leave them anywhere?
79. If people don't want to do the same things as you: a) does it make you mad, or b) do you just do what they want to do?

### Interpretation of Factors

In our earlier reports, we offered some general interpretations of questionnaire factors at the middle-childhood level. We have refrained, however, from any extensive interpretation on the basis of item content alone. But we have now assembled personality scales for twelve factors on a purely statistical basis. We can offer sound mathematical arguments for the utility of these scales, for they have been designed to cover a broad area of personality with maximal economy. The actual utility of these scales, on the other hand, now depends on our understanding of their psychological referents.

Apart from manifest item content, there are two chief avenues by which we can hope to gain some insight into the factors that we are measuring: a) we can determine the way in which questionnaire scores covary with other kinds of personality measures, and b) we can link this questionnaire instrument with those developed for other age levels by using a sample at an intermediate age level. The first type of comparison can be conveniently made, since we also have teacher-rating, parent-rating, and objective-test data for our Decatur sample. The analyses of these data have been reported elsewhere (1, 3, 5). In another article (2), we shall consider interrelations among factor scores derived from all of our various media of measurement and describe a common factor analysis of these scores.

For the other type of comparison, it is convenient to relate the ESPQ to the High School Personality Questionnaire (HSPQ). The latter test, which is described elsewhere (6), was devised for the 12-to-16-year range. It contains 14 scales which have been identified in terms of factors that have appeared in studies of adults and pre-adolescents. Our plan was to give the ESPQ and the HSPQ to a sample at about the nine- and ten-year level. It must be recognized, of course, that some of the ESPQ will seem a bit "childish" to such a group, while portions of the HSPQ will be incomprehensible to them. Some resultant distortion of correlations between the two tests is to be expected. No other procedure, however, permits such a direct comparison of questionnaire factors at the two age levels from which the two tests were derived.

We can report findings for two independent samples.* The first sample consisted of two fourth-grade and two fifth-grade classes in Terre Haute, Indiana. There were 46 boys and 59 girls, making a total of 105. In age, they ranged from nine years and five months to eleven years and eleven months, with a mean of ten years and five months. These children received all items of the ESPQ, except for factor 13, at one session and Form A of the HSPQ at another session. The product-moment correlations between the ESPQ total scores and Form A scores for the HSPQ are shown in Table I.

A second sample was secured in Kent, Ohio. It consisted of one third-grade class and three fourth-grade classes. There were 43 boys and 49 girls, or a total of 92 children, ranging in age from eight years and five months to ten years and eleven months, with a mean of nine years and nine months. For this sample, we were able to obtain total scores, based on both forms, for all scales of each test. The lower age level of the Kent sample is conducive to some loss of reliability and validity for the HSPQ. Any such loss, however, was probably offset by the increased stability afforded by the use of both forms of this test. In fact, we might justifiably attach more weight to the Kent data than to the Terre Haute data. The product-moment correlations between ESPQ and HSPQ scores for the Kent sample are shown in Table II.

In each of our tables, there are several rows and columns in which more than one substantial correlation can be seen. This suggests the definite possibility of age changes in personality structure. At either age level (six-to-seven or twelve-to-sixteen), there could be factors intermediate between factors appearing at the other level. There could be processes of factor convergence, whereby eleven-year factors are discernible at the six-year level only in the second-order realm, or of factor divergence, whereby six-year factors become second-order at the eleven-year level. Our present data, however, do not provide a sufficient basis for drawing any such conclusions. The picture is complicated by the approximate character of factor scores estimated by the present methods and by the fact that, at every age level thus far studied, there are some factors that inevitably tend to "cooperate", or share high-loading variables. In the construction of item scales for cooperative factors, a complete suppression of irrelevant factor variance is often not possible. For this and other reasons, the intercorrelations within one set of factor scales will not faithfully represent the actual intercorrelations among the factors. Some less severe distortion of correlations with outside measures is to be expected. While possibilities for developmental shifts in factor structure thus exist, they must be examined in the light of other evidence than that presented here.

In any case, some of the ESPQ factors can be identified with some certainty if both the item content and the HSPQ correlations are carefully weighed together. Any given correlation, of course, must be evaluated in the light of all other correlations obtained for its component variables.

Factor 1 of the ESPQ shows some affinity with factors A, E, and I of the HSPQ. Since E and I seem to be represented elsewhere, factor A looms as the best match. Such an interpretation appears to be consistent with the content of the items. The items, however, stress primarily a form of socialization incidental to adjustment in a school environment. This suggests something akin to factor K, or Socialized Morale, which has appeared only in rating data in previous studies of older groups. In its present form, then, factor 1 may be identified as A- (Schizothymia vs. Cyclothymia) or as K- (Dislike of Education vs. Socialized Morale).

In factor 2, we see a pattern of irritability and subjective distress, which suggests O (Anxious Depression) or $Q_4$ (Nervous Tension). At the same time, there is much in the item content that suggests the pattern of overreaction and impulsiveness of factor D (Excitability or Infantile Emotionality). This elusive factor does not reveal itself clearly in questionnaire research with older groups, but present findings are consistent with Excitability as an interpretation of our present factor.

Like factor 1, factor 3 seems to involve attitudes and adjustment with respect to school. More generally, it seems to relate to an acceptance of adult authority and a successful incorporation of adult standards of conduct. The correlations are consistent with an interpretation of this factor as G (Superego Strength).

Factor 4 seems to combine emotional sensitivity, timidity, and aesthetic interest in a way characteristic of factor I (Premsia or Emotional Sensitivity), but the opposing content is strongly

---

*For the procurement of these samples, the authors are indebted to Dr. Rutherford B. Porter and Dr. Edna R. Oswalt and to the cooperating school personnel of Terre Haute, Indiana, and Kent, Ohio.

TABLE I

CORRELATIONS BETWEEN ESPQ AND HSPQ FOR FIRST SAMPLE
(With Decimal Points Omitted)

| ESPQ total scores | HSPQ Form A Scores | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J | O | $Q_2$ | $Q_3$ | $Q_4$ |
| 1 | -34 | 16 | 34 | 12 | 40 | -09 | -28 | 21 | -65 | -26 | -23 | 12 | -26 | -20 |
| 2 | -21 | 02 | -12 | 30 | 19 | -14 | -11 | -10 | -06 | -09 | 20 | -06 | -19 | 19 |
| 3 | 25 | 06 | 02 | -22 | -57 | 08 | 34 | 17 | 34 | 20 | -02 | 14 | 26 | -02 |
| 4 | 37 | -19 | -29 | -12 | -41 | 17 | 24 | -07 | 63 | 34 | 13 | -21 | 26 | 16 |
| 5 | 21 | -08 | -05 | -26 | -40 | 12 | 27 | 23 | 19 | 29 | -08 | 09 | 15 | -10 |
| 6 | 27 | -01 | 21 | -47 | -45 | 23 | 31 | 37 | -00 | 13 | -36 | 13 | 44 | -23 |
| 7 | -11 | 08 | 20 | 14 | 30 | -01 | -23 | 14 | -46 | -19 | -06 | -01 | -28 | -15 |
| 8 | -08 | -03 | 20 | -03 | 15 | -12 | -27 | 13 | -42 | -17 | -21 | -06 | -28 | -22 |
| 9 | -19 | 06 | -36 | 38 | 14 | -20 | 05 | -30 | 22 | -00 | 41 | 02 | -06 | 31 |
| 10 | -23 | 16 | 29 | 06 | 26 | -12 | -19 | 19 | -57 | -35 | -19 | 09 | -18 | -26 |
| 11 | 12 | 10 | 04 | 02 | -22 | 13 | 24 | 13 | 20 | 25 | -07 | -07 | 14 | -12 |
| 12 | 34 | -18 | -07 | -36 | -54 | 03 | 39 | 20 | 33 | 37 | -07 | 15 | 31 | -12 |

### TABLE II

### CORRELATIONS BETWEEN ESPQ AND HSPQ FOR SECOND SAMPLE

(With Decimal Points Omitted)

| ESPQ total scores | HSPQ Total Scores | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J | O | $Q_2$ | $Q_3$ | $Q_4$ |
| 1 | -54 | 00 | -14 | 18 | 44 | -15 | -35 | -31 | -50 | -24 | 23 | -05 | -31 | 21 |
| 2 | -49 | -09 | -58 | 57 | 49 | 09 | -39 | -64 | 08 | -01 | 64 | -20 | -30 | 67 |
| 3 | 38 | 06 | 21 | -35 | -44 | -19 | 48 | 40 | 18 | 23 | -30 | 15 | 31 | -43 |
| 4 | 57 | 03 | 26 | -44 | -57 | 06 | 43 | 48 | 49 | 29 | -33 | 20 | 39 | -27 |
| 5 | 39 | 09 | 30 | -35 | -38 | -18 | 41 | 43 | 19 | 09 | -40 | 17 | 21 | -29 |
| 6 | 58 | 28 | 49 | -59 | -53 | -05 | 46 | 50 | 20 | 25 | -45 | 29 | 45 | -60 |
| 7 | -10 | -10 | -11 | 36 | 43 | 38 | -32 | -17 | -34 | -48 | -04 | -13 | -26 | 28 |
| 8 | -37 | -06 | -10 | 14 | 37 | -08 | -15 | -19 | -32 | -29 | 16 | -11 | -18 | 09 |
| 9 | -45 | -17 | -44 | 24 | 28 | -18 | -09 | -31 | -04 | 01 | 21 | -05 | -21 | 25 |
| 10 | -33 | -03 | -05 | 15 | 37 | -06 | -26 | -14 | -57 | -26 | 12 | -05 | -24 | 05 |
| 11 | 04 | -01 | -04 | -16 | -10 | 03 | 04 | -08 | 07 | -02 | 01 | 03 | -01 | -10 |
| 12 | 49 | 28 | 45 | -44 | -46 | 07 | 44 | 50 | 23 | 21 | -45 | 26 | 27 | -50 |
| 13 | 09 | 32 | 02 | -11 | 08 | 07 | -04 | -05 | -08 | 01 | 00 | 17 | 05 | -16 |

suggestive of factor E (Ascendance vs. Submissiveness). The correlations support interpretations in terms of factor A, factor E, and factor I. Factors A and I are probably both better represented by other factors in the present set. Further evidence may clearly identify the present factor as E- (Submissiveness vs. Ascendance).

The most prominent feature of factor 5 seems to be a certain dependence on parents and teacher, with a conformity to their wishes. The correlations are generally consistent with this content, but they do not point to any single clear match with a recognized factor. The two matrices agree in relating this factor to A, D-, E-, G, and H, but each of these seems to be better represented by another factor in the set.

Factor 6 shows emotional control and tolerance of frustration at the high end and a more ready loss of composure at the low end. The correlations most strongly suggest identification of this factor as D- (Phlegmatic Frustration Tolerance vs. Infantile Emotionality or Excitability), but the content is also quite consistent with O- (Confidence vs. Anxious Depression).

At the high pole of factor 7, we see an aggressive self-assertion which is consistent with the pattern of factor E (Ascendance vs. Submissiveness) at lower age levels. Another possible interpretation is F (Surgency vs. Desurgency), which, unlike E, is apparently not represented elsewhere in the present set of factors. Factor 8 has no unambiguous match, though the correlations indicate some relationship to A, E, G, I, J, and $Q_3$. The prevailing theme running through the items of this factor might be characterized as "motor expressiveness vs. socialized restraint."

In the present correlation data, factor 9 aligns best with C. The item content accords generally with such a match, though it is a little more strongly suggestive of A or H. The high scorer appears to combine certain schizothyme qualities of quietness, seclusiveness, and sullenness. A and H are supported by the correlations, though less strongly than C. Since A appears more clearly elsewhere, factor 9 may be best identified as H- (Withdrawn Schizothymia vs. Adventurous Cyclothymia) or C- (Poor integration vs. Ego Strength).

Factor 10 correlates best with I, and the content is clearly consistent with this identification. At its positive pole, there is a certain boisterousness and toughness. At the negative pole, we find fearfulness and timidity. We may tentatively identify this factor as I- (Rigid, Tough Poise vs. Sensitive Emotionality).

Factor 11 shows no marked relationship to any HSPQ factor, and no clear overall pattern is readily discernible in the items. It seems best to refrain from interpretation until further data are available. Factor 12 presents a pattern of friendly sociability and calmness at the high end. The aggressiveness at the low end of the scale accounts for a high negative association with E, but the overall pattern at the high end is not that of submissiveness. It is more nearly one of cyclothymia. Both A and H, however, appear to be represented elsewhere. $Q_4$ (Low Nervous Tension vs. High Nervous Tension) is another possible interpretation.

Factor 13 was not used in the Terre Haute study. In the Kent study, it correlates appreciably only with factor B (Intelligence). This unambiguous finding actually exceeds expectations, for a correlation even as high as that obtained (+.32) is surprising in view of the fact that the ESPQ and HSPQ intelligence scales are severely restricted at opposite ends of the range in our sample. At the intermediate age level of the Kent sample, the ESPQ fails to discriminate at the high end of the range, while the HSPQ fails to discriminate at the low end.

## Summary

Two equivalent forms were constructed for a factored personality questionnaire suitable for group administration to children in the six to eight year range. Each form contains 12 personality scales derived from two earlier factor-analytic studies. A thirteenth scale for intelligence was added to each form. Scores on the questionnaire were correlated with those for a questionnaire designed for older children for two independent samples. Inspection of the correlation data and of the items themselves permits satisfactory identification of most of the factors. There are indications that the test taps each of the following known factors:

A. Cyclothymia vs. Schizothymia
B. General Intelligence
C. Ego Strength vs. Poor Integration
D. Infantile Emotionality
E. Ascendance vs. Submissiveness
F. Surgency vs. Desurgency
G. Superego Strength vs. Dependent Immaturity
H. Adventurous Cyclothymia vs. Withdrawn Schizothymia
I. Emotional Sensitivity vs. Toughness
K. Socialized Morale vs. Dislike of Education
O. Anxious Depression
$Q_4$. Nervous Tension

In general, it seems inadvisable to accept the factor identifications which we have indicated as final. There are some departures from the factors found among adults that accord with other findings on the composition of known factors in childhood. There are a few factors that do not align clearly with any known factors. Further refinement of the scales should reveal whether these represent real structural novelties at the middle-childhood level or simply poor approximations to identifiable factors. A study dealing with the relationships among questionnaire, rating, and objective-test factors (2) may provide further clarification.

## REFERENCES

1. Cattell, R. B., and Coan, R. W. "Child Personality Structure as Revealed in Teacher's Behavior Ratings," Journal of Clinical Psychology, XIII (1957), pp. 315-327.
2. Cattell. R. B., and Coan, R. W. A Comparison of Factors Derived from Different Realms of Measurement in Middle Childhood (in preparation).
3. Cattell, R. B., and Coan, R. W. Objective-Test Assessment of the Primary Personality Dimensions in Middle Childhood, British Journal of Psychology (in press).
4. Cattell, R. B., and Coan, R. W. "Personality Dimensions in the Questionnaire Responses of Six and Seven-Year-Olds," British Journal of Educational Psychology, XXVIII (1958), pp. 232-42.
5. Cattell, R. B., and Coan, R. W. "Personality Factors in Middle Childhood as Revealed in Parents' Ratings," Child Development, XXVIII (1957), pp. 439-58.
6. Cattell, R. B., Coan, R. W., and Beloff, Halla. A Re-examination of Personality Questionnaire Factors in Late Childhood and Development of the HSPQ (in preparation).
7. Coan, R. W., and Cattell, R. B. "Reproducible Personality Factors in Middle Childhood," Journal of Clinical Psychology, XIV (1958) pp. 339-45.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 2, December 1959)

# THE RELATIVE VALIDITY OF VERBAL VS: ILLUSTRATION ITEMS IN ACHIEVEMENT TYPE TESTS FOR LOW VERBAL ABILITY PERSONNEL*

PETER F. MERENDA
Walter V. Clarke Associates, Inc.**
CHARLES J. MACALUSO
U. S. Naval Examining Center

## Problem

NON-VERBAL group tests have received wide attention in psychological testing programs although their use in this country has been rather limited. The most extensive use of the non-verbal group test has been in the military services where the problem arises of processing and classifying great numbers of illiterates and linguistically-handicapped persons. Basically these non-verbal tests have been of the general classification and academic aptitude type and have been used primarily for screening and classificatory purposes. Their development and use has stemmed from the premise that personnel with linguistic and reading handicaps are penalized on a test score basis on the attribute ostensibly being measured. However, the development of these tests has been seriously handicapped by the difficulty of producing completely non-verbal items. Except for the Army Beta Examination (8) and the Semantic Test of Intelligence (5) in which the directions are given completely in pantomime, mostly so-called non-verbal tests are merely partly non-verbal tests. Furthermore, they are difficult to administer, expensive to produce, and the individual items in these tests are generally too easy for personnel of superior ability. Probably for these and other apparent reasons the use of non-verbal group tests has been quite limited. In any event, even their limited use has been further restricted to the area of classification testing and these tests have not found application in the field of achievement testing where the performance test generally replaces the paper-and-pencil test when the latter is considered inappropriate.

In the U. S. Navy, a system of competitive service-wide examinations for the selection and advancement of petty officers has been established. All candidates for petty officer rates in the 60-odd naval enlisted occupations who have met minimum requirements and who have been recommended by their superior officers for advancement to the next higher pay grade level compete on these examinations which are administered semi-annually throughout the naval service. These examinations consist of 150 multiple-choice items designed to sample the professional and military knowledges and skills required for successful job performance in the rate being sought. They are further designed as highly discriminating selection tests since only those candidates are allowed to compete who have been selected by their commanding officers and who have been certified by them principally on the basis of training and on-the-job performance to be at least minimumly qualified for advancement.

In order to provide for a universally fair and equitable system, all candidates for advancement are required to compete with their contemporaries on the identical standard. The principal component of this selection criterion is the standardized paper-and-pencil achievement type test described above.

Among the 60-odd occupations which the U. S. Navy maintains in its petty officer structure at four pay grade levels, there is considerable variation both between and within groups in verbal and linguistic ability. In the more highly specialized and technical ratings such as Electronics Technician, Air Controlman, and Aerographer, although there is reasonably high variance in this attribute within the groups, the overall ability is, on the average, generally high and does not appear to constitute much of a problem in terms of its contribution to error variance on the advancement test score variable. However, in the less technical service ratings such as Steward's Mate, Ships Serviceman, and Commissaryman, the relatively high variance coupled with a generally low overall verbal ability has led to the formulation of the hypothesis that completely verbal items in the advancement tests tend to reduce the validity of the measurement of job proficiency for members of these low verbal ability groups. The study reported in this paper was designed to test this hypothesis.

---

* The opinions and assertions expressed in this paper are solely those of the authors, and do not necessarily reflect the official views of the Navy Department.

**Complete address: Walter V. Clarke Associates, 324 Waterman Ave., East Providence 14, R. I.

## Procedures and Results

It was decided to conduct this research in the rating of Steward's Mate at the level of third class petty officer candidates since the Steward's Branch in the Navy represents the lowest level of verbal ability. In addition to the generally low verbal ability of personnel in this rating, there exists the additional feature of linguistic-handicap (bilingualism) caused by the employment of great numbers of native-born and native-educated Filipinos.

For the purposes of this study verbal ability is defined and measured by performance on the General Classification Test of the Navy Basic Test Battery. Factor analysis (9) and correlational (7) studies have shown this test to possess an exceptionally high verbal factor loading. Scores on this test are reported as Navy Standard Scores ($\overline{X} = 50$; $\sigma = 10$). From samples drawn from among candidates for advancement to petty officer, third class in 52 naval occupations in February 1955, Steward's Mates proved to represent the significantly lowest level on verbal ability ($\overline{X} = 35.70$; $\sigma = 8.40$). The next higher level was reached by Gunner's Mates ($\overline{X} = 40.47$; $\sigma = 7.77$) (C.R. = 5.58).

A team composed of a subject-matter specialist (Chief Steward's Mate employed at the Naval Examining Center as item writer for Steward's Branch Examinations), a professional test technician (civilian educational specialist normally assigned to the task of reviewing and editing items in Steward's Branch Examinations), and a commissioned officer of the Supply Corps who normally reviews items for Steward's Branch Examinations from the standpoint of technical accuracy worked on the project of developing the experimental examination forms.

In developing these test forms, items were either drawn from item card files or were constructed as original items. The criterion standards which items were required to meet in order to be considered acceptable for inclusion in this examination were:

1) They must have sampled the petty officer requirements (1) of Steward's Mates particularly at the third class level; 2) for items drawn from the files, they must have been of appropriate difficulty level ($p > .10 < .90$) and, also, of positive and significant discrimination ($\overline{D} > +0.4$) (3); and 3) the material upon which each item was to be based must have been of such a nature that it would lend itself equally well to presentation both in completely verbal form and in picture form (illustration item).

Accordingly, a sufficient number of items were either selected or constructed to meet all three of the above criteria. Each item was developed in two forms: 1) completely verbal form—verbally presented stems and options (four choices), and 2) illustration form—verbal stem with options presented in picture form. In both forms the use of language was kept to a minimum, and to the extent possible the words employed were kept in the vernacular.

Two test forms, each consisting of the same fifty items were developed. In form II, the form composed of the picture items, the order of presentation was reverse from that of form I (verbal form). In all other respects, the items were identical, except for the mode of presentation. The same explicit and careful directions and general instructions were presented on the cover page of both forms, and both were reproduced by commercial offset printing methods. In the printing of the final forms, however, the illustrated form was reproduced about twice the size of the verbal form in an attempt to minimize any confusion which migh arise on the part of the subjects to be tested by the lack of detail in the pictures which sometimes results from the reduction in size from original to final form by this method of printing.

The experimental tests were administered in September 1955, to a total of 111 Stewardsmen representing nine different fleet units and stations who were assembled at the Naval Air Station, North Island, San Diego, California. Prior to administering the tests, the group of 111 subjects was divided into two subdivisions ($N_1 = 67$ and $N_2 = 44$). Sub-group one was administered the verbal test form first and the illustrated test form second. For sub-group two, the above procedure was reversed. This process was employed in order to minimize the influence of "practice effect." Originally, it was planned to administer the forms alternately to the entire group, but physical facilities and other factors made it necessary to divide the groups in this manner and in these numbers, and to administer the tests in two separate testing areas.

## Analysis of the Data with Interpretations

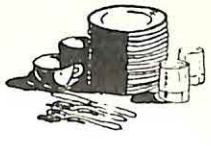
Total Sample—For the 111 subjects, means and standard deviations were calculated for scores obtained on each test form; and then employing the difference scores, the difference between these means was tested for significance. These results

SAMPLE QUESTION

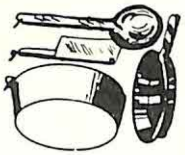
SAMPLE ANSWER FORM

1. Which is NOT kept in the officers' galley?

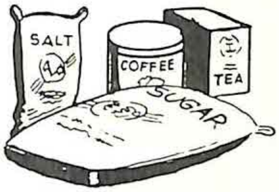
1.



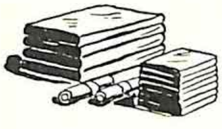
3.

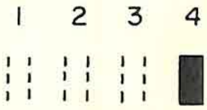


2.



4.





1 2 3 4

are presented in Table I. Only a very slight difference was found between the mean scores obtained by the total sample on the verbal and illustration forms of this test. Statistically, this difference borders on significance at the five percent level of confidence. However, it does appear that variation of the mode of presentation of items does not greatly influence the total test scores.

American Group vs. Filipino Group—In order to discover whether or not the mode of presentation of items might show different effects for the more seriously verbal-handicapped than for the lesser linguistically-handicapped, the total group was divided into American-born and Philippine-born subjects.

The differences between mean scores on both test forms obtained by originally English speaking subjects and originally Filipino speaking subjects proved to be statistically significant. A further inspection of the data in Table II reveals that the mean scores and variances obtained by these two groups on the verbal and illustration forms are practically the same, and that they are very close to the total sample values. Here again, it is shown that mode of item presentation does not influence final test scores.

Higher Verbal Group vs. Lower Verbal Group—A further comparison was made between the scores obtained on each form by subjects of higher vs. lower verbal ability as measured by the Navy General Classification Test. Distribution of GCT scores of the total sample (N = 107)* yielded a mean of 36.67 and a standard deviation of 6.91. $Q_1$ and $Q_3$ values were determined and were found to be, respectively, 30.8 and 40.4. The high and low groups were selected as close to these points as possible (31 and 40). Due to skewness in the distribution and rounding-off procedures, the resulting sub-group N's were $N_{Q_3} = 37$ and $N_{Q_1} = 27$.

The data in Table III again demonstrate the lack of influence of mode of presentation within groups of both higher verbal ability and lower verbal ability. It is noted, however, as expected, that the test score means on the respective forms of the subjects within the higher verbal ability group are significantly higher than those of the lower verbal ability group. Here, again, the hypothesis of no difference must be accepted for both groups.

Study of "Practice Effect"—Since the subjects included in this sample were scheduled to take both forms of this experimental test in one sitting, the situation required that "practice effect" be controlled to the highest degree possible. Unfortunately, however, circumstances forced the administrators to alter their original plans of administering the forms alternately to all subjects seated in a single testing room. Restriction of facilities both within and outside the testing room necessitated the splitting up of the total sample into two separate groups seated in two distinct testing areas. The sub-groups were randomly selected and again, due to physical restrictions it was impossible to equate the groups in size. Group I numbered 67 and was administered the verbal form first, then the illustration form. Group II was composed of 44 subjects, and for them the administration procedure was reversed. The groups were kept intact and were separated during a short break period between administrations. Both were told that they would be taking two examinations on the same subject but in different forms. Neither group was told that the items contained in one form were identical to those in the other form. No strict time limits were imposed on either group, but the administration time for both groups was approximately equal. The GCT scores were studied for subjects who took the verbal form first and those who took the illustration form first.

---

*There were four subjects for whom GCT scores were not available.

TABLE I

DISTRIBUTION STATISTICS AND SIGNIFICANCE VALUES FOR SPECIAL STEWARD EXAMINATION (VERBAL FORM AND ILLUSTRATION FORM)

| Test Form | N (No. of pairs) | Raw Score Mean | Raw Score SD | SD of Diff. Score | CR | P |
|---|---|---|---|---|---|---|
| Verbal | 111 | 28.22 | 4.32 | 4.00 | 2.05 | <.05 |
| Illustration | | 29.00 | 4.07 | | | |

TABLE II

DISTRIBUTION STATISTICS AND SIGNIFICANCE VALUES FOR AMERICAN-BORN AND PHILIPPINE-BORN SUBJECTS

| Test Form | Group | N | Raw Score Mean | Raw Score SD | CR | P |
|---|---|---|---|---|---|---|
| Verbal | American | 67 | 28.81 | 4.07 | 1.81 | > .05 |
| | Filipino | 44 | 27.30 | 4.41 | | |
| Illustration | American | 67 | 29.16 | 4.29 | 0.56 | > .05 |
| | Filipino | 44 | 28.75 | 3.75 | | |

## TABLE III

DISTRIBUTION STATISTICS AND SIGNIFICANCE VALUES FOR VERBAL AND ILLUSTRATION FORM TEST SCORES OF HIGHER AND LOWER VERBAL ABILITY GROUPS

| Verbal Ability Group | Test Form | N (No. of pairs) | Raw Score Mean | Raw Score SD | SD of Diff. Score | CR | P |
|---|---|---|---|---|---|---|---|
| Higher | Verbal | 37 | 29.59 | 4.36 | 1.08 | 0.4129 | > .05 |
| | Illustration | | 30.32 | 3.77 | | | |
| Lower | Verbal | 27 | 26.70 | 3.78 | 4.23 | 0.3690 | > .05 |
| | Illustration | | 26.67 | 3.55 | | | |

## TABLE IV

DISTRIBUTION STATISTICS AND SIGNIFICANCE VALUES OF GENERAL CLASSIFICATION TEST SCORES OF SAMPLE TAKING VERBAL FORM FIRST vs. SAMPLE TAKING ILLUSTRATION FORM FIRST

| Group | N | GCT Mean | SD | CR | P |
|---|---|---|---|---|---|
| Verbal Form 1st | 64* | 36.70 | 6.74 | 0.0359 | > .05 |
| Illustration Form 1st | 43** | 36.65 | 7.17 | | |

* 3 subjects for whome GCT scores were not available.
**1 subject for whom GCT score was not available.

The practically identical means ($M_D = 0.05$) show that the random selection of subjects to Groups I and II prior to the administration of the tests was indeed effective in producing two groups equated with respect to general verbal ability.

To test for the presence of any "practice effect" within groups, means and standard deviations were calculated on each test for those subjects who took the verbal form first and those who took the illustration form first, and the differences between means were tested for significance.

The data in Table V reveal the existence of some "practice effect." It will be noted that for each sub-group there was a statistically significant increase in the average scores obtained on the test of the second administration. This finding is as expected since it is practically impossible to completely control this effect as a source of error variance. It is only hoped that this effect will be kept at a minimum, and in this light the fact that the differences between mean scores, although statistically significant, were not too great was encouraging.

A further study of the "practice effect" was made by comparing the results between groups on the two forms of the test. Means and standard deviations of verbal form scores and illustration form scores were compared between those yielded by the group taking the verbal form first and those taking the verbal form second.

The data in Table VI reveal evidence of the counter-balance influence on "practice effect" by the employment of sub-samples alternately administered the two test forms. Although the data in Table V show evidence of the existence of "practice effect" the fact as revealed in Table VI that neither form was more difficult or easier than the other on both administrations and that the differences in mean performance were in direct proportion to the order in which taken demonstrates the efficiency of the administration procedures in minimizing the influence of "practice effect." The data in Table VI also show that there was less change in scores from one administration to the other with the illustration form ($P > .05$) than with the verbal form ($P < .01$).

## Reliability and Validity of the Test Forms

*Reliability*—The reliability of these forms was determined by the Rulon (5) Method applied to parallel forms and correlated samples. The formula employed for determining this reliability coefficient is identical to the formula for deriving the correlation coefficient between a set of difference scores.

$$r_{1I} = \frac{\sigma 1^2 + \sigma 2^2 - \sigma d^2}{2\sigma_1 \sigma_2}$$

For these tests the reliability coefficient was found to be .55. Since the experimental forms were each composed of only 50 items whereas the standard examinations are comprised of 150 items, an estimate of the total test reliability was computed according to Guilliksen (2), and was found to be equal to .79. This is evidence of reasonably high test reliability.

For each test form a K-R #20 reliability estimate was obtained. This formula yielded an internal consistency coefficient of .47 for the illustration form (50 items) and .52 for the verbal form (50 items). These statistics transformed to Z's and tested for significance proved to be homogeneous ($\chi^2 = 0.19$); thus yielding further evidence of equal reliability for both forms. The lower average value of the Kuder-Richardson coefficient as compared to the Rulon coefficient is to be expected since dispersion of item difficulties and item heterogeneity were present in the tests. These are conditions under which the K-R #20 coefficient is attenuated.

*Validity*—An external criterion was available for establishing the predictive ability of these tests. The validity criterion was the Advancement in Rating Examination standard score obtained by 47 subjects of the total sample of 111 on the August 1955 Navy-wide competitive Steward's Mate examinations. Both verbal and illustration form scores for this sample were individually correlated with the criterion scores. In each case, the resulting correlation coefficient was .53. This value was corrected for increase in length of the total test ($K_t = 150$) by a method described by Gulliksen (2: 89) and the final validity coefficient for each test form was found to be .63.

## Item Analysis Data with Interpretations

*Total Sample*—The responses selected by the subjects to each of the parallel items were subjected to an extensive item analysis treatment. The first treatment involved the separating of the total sample into 25% high and 25% low on each test form, and computing difficulty level estimates and discrimination indices according to Lawshe (3) for each of the parallel items. These item analysis data for each set of parallel were then compared by visual inspection for apparent real differences. Since these data represented correlated proportions (repeated measurements on the same subjects) the conventional statistical test of significance was not appropriate, and therefore could not be applied to the data in this form. A rough estimate of real differences between parallel items based upon this type of subjective judgment revealed that less than twenty-five percent of items appeared to show differences in difficulty level and power of discrimination when presented in different mode. These item analysis data showed that the great majority of the items in both forms proved to be high positive in discrimination power

### TABLE V

### DISTRIBUTION STATISTICS AND SIGNIFICANCE VALUES OF VERBAL AND ILLUSTRATION FORMS SCORES FOR SUB-GROUPS TAKING ALTERNATE FORMS FIRST

| Group | Test Form | N (No. of pairs) | Raw Score Mean | Raw Score SD | SD of Diff. Score | CR | P |
|---|---|---|---|---|---|---|---|
| Verbal Form First | Verbal | 67 | 26.93 | 4.09 | 3.27 | 6.38 | <.001 |
| | Illustration | | 29.45 | 4.07 | | | |
| Illustration Form First | Verbal | 44 | 30.18 | 4.47 | 3.42 | 3.59 | <.001 |
| | Illustration | | 28.32 | 4.02 | | | |

TABLE VI

DISTRIBUTION STATISTICS AND SIGNIFICANCE VALUES OF VERBAL vs. ILLUSTRATION FORM SCORES OBTAINED BY SUBJECTS TAKING ALTERNATE FORMS FIRST

| Test Form | Group | N | Raw Score Mean | Raw Score SD | CR | P |
|---|---|---|---|---|---|---|
| Verbal | Verbal First | 67 | 26.93 | 4.09 | 3.00 | <.01 |
| | Illustrated First | 44 | 30.18 | 4.47 | | |
| | Total | 111 | 28.22 | 4.32 | | |
| Illustration | Verbal First | 67 | 29.45 | 4.07 | 1.44 | > .05 |
| | Illustrated First | 44 | 28.32 | 4.02 | | |
| | Total | 111 | 29.00 | 4.07 | | |

and to possess high internal consistency.

To test the statistical significance of the difference between difficulty levels of the parallel items, a method described by McNemar (4:55-59) was employed. The total sample of 111 cases was randomly reduced in size to 100 in order to facilitate the computation of proportion values, and the proportions passing and failing each parallel item were analyzed.

Of the 50 sets of item pairs (verbal vs. illustration presentation) a total of 35 yielded differences in difficulty level which were not statistically significant. Of the 15 pairs of items in which statistically significant differences occurred, nine proved to be less difficult in illustration form, and six were less difficult in verbal form. One of the nine item pairs which appeared to be less difficult in illustration form (C. R. = 2.12) proved to be almost as easy in verbal form ($P_I = .99$, $P_V = .93$). For these fifteen items, the following specific findings are reported:

1. Items in which fine discriminations and distinctions are required to be made appear to increase in difficulty when presented in illustration form. For example, in both item sets in which the examinee is required to discriminate between a dessert spoon and a teaspoon, a significant increase in difficulty of the illustration form was noted.
2. In items which require a close attention to detail there also appears to be a corresponding increase of difficulty level in illustration form. For example, in item sets in which the examinee is required to pay close heed to details which are somewhat confounded in the reproduction of the illustration, there followed a substantial increase in the difficulty level of the illustrated item.
3. In items in which either clearly descriptive action or identification of things is involved, there appears to be a decrease in the difficulty level of the illustrated item. This was evident in eight of the fifteen item sets which produced statistically significant differences.

## Summary and Conclusions

To test the hypothesis that completely verbal items in achievement type tests tend to reduce the validity of measurement for members of low verbal ability groups, two experimental test forms were developed and administered to a sample of 111 Navy Stewardsmen, 44 of whom were native-born Filipinos and 67 of whom were native-born American Negroes. Both test forms contained the identical items, each sampling knowledges and skills required for promotion to Steward's Mate, Third Class. In one form the mode of item presentation was completely verbal, whereas in the other form the items were in illustration form (verbal stems, illustrated options).

Both test forms were administered to the same subjects, and comparisons of performance were made both on a total score basis and for individual pairs of parallel items. Both experimental tests proved to be parallel forms. Reliabilities and validities of each form were also identical. Of the 50 item sets (pairs) 35 proved to be of equal difficulty level; and of the 15 which showed statistically significant differences, nine were less difficult in illustration form, and six were more difficult. Also no significant differences were found to exist between the performance on both test forms of American-born and Philippine-born subjects; and between subjects of higher and lower verbal abilities.

From the findings obtained in this study, the following conclusions appear to be justified:

1. Varying the mode of presentation (from completely verbal to partially verbal form) of achievement type test items does not appreciably affect the predictive validity of the measurement when applied to members of the low verbal ability group studied.
2. For the major portion of the material upon which this type of measurement is based, presentation in either form produces reasonably the same results.
3. The required increase in time, effort, and nonetary costs involved in the development and construction of large numbers of illustration items would not be justified.

## REFERENCES

1. Bureau of Naval Personnel, *Manual of Qualifications for Advancement in Rating, NavPers 18068* (Washington, D.C.: Navy Department 1952), pp. 12-15.
2. Guilliksen, H. *Theory of Mental Tests* (New York: John Wiley and Sons, 1950), pp. 78-89.
3. Lawshe, C. H. *Principles of Personnel Testing* (New York: McGraw-Hill, 1948), pp. 183-91.
4. McNemar, Q. *Psychological Statistics* (New York: John Wiley and Sons, 1955), pp. 56-59.
5. Rulon, P. J. "A Simplified Procedure for Determining the Reliability of a Test by Split-Halves," *Harvard Educational Review*, XXX (1939), pp. 99-103.
6. Rulon, P. J. "A Semantic Test of Intelligence," *Proceedings, 1952 Invitational Conference on Testing Problems* (Princeton, N. J.: Educational Testing Service, 1952), pp. 84-92.
7. Stuit, D. B. (Ed.), *Personnel Research and Test Development in the Bureau of Naval Personnel* (Princeton, N. J.: Princeton University Press, 1947), p. 72.
8. Yerkes, R. M. (Ed.), "Psychological Examining in the United States Army," *Memoirs*,

(Washington, D.C.: National Academy of Science, 1921), p. 890.
9. Zachert, V. Factor Analysis of the Army, Navy, and Air Force Classification Batteries, Research Bulletin 52-12 (San Antonio, Texas: Human Resources Research Center, Lackland Air Force Base, March 1952), p. 6.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 2, December 1959)

# AURAL PERCEPTIONS OF AUTHORITARIANS IN DIFFERENT CULTURAL SITUATIONS

JACK R. FRYMIER
Auburn University
Auburn, Alabama

## The Problem

IN THE FIELD of perceptual psychology, it is sometimes held that behavior is a function of perception (15).

Authoritarian behavior has been the subject of much research in recent years, and certain psychologists hold that those persons classed as "authoritarians" are less able to perceive reality, tolerate ambiguity, resist suggestion, and assert their independence than are "nonauthoritarians." Accordingly, if behavior is a function of perception, then these behavior characteristics of authoritarians may be products of the way they perceive. One purpose of the present study was to explore this possibility.

Further, there is general feeling that an individual's cultural background exerts considerable influence on his personality development. If this view is correct, interest attaches to the way in which the culture exerts its influence. Is it that a particular cultural situation tends to develop in the individual certain perceptions, and in this manner influences the very fabric of his personality? In the present study, groups of persons from two different cultural backgrounds were included in order to obtain some evidence on this particular problem.

Maslow (13) first outlined the authoritarian character in 1943, listing, among other things, some of the behavioral characteristics which were examined in this particular study. Adorno (1) and his colleagues devoted considerable time and space reporting their studies of authoritarian personalities, out of which emerged the F (Fascist) scale, a measure of anti-democratic potential. Christie and Jahoda (3) examined the entire research procedure involved in the development of the F scale and seriously questioned some of the techniques and conclusions of Adorno and his colleagues. Titus and Hollander (17) reviewed over sixty studies which used the F scale in psychological research from 1950 to 1955 and concluded that the concepts of "personality" and "syndrome" were without meaning. They further concluded that the use of the F scale as a predictive instrument was still questionable. Masling (12) reviewed four studies of authoritarianism and concluded that the characterization of the authoritarian had been overdrawn. Drucker (5) studied college students and student nurses and reported that conformity was related to authoritarianism. Koontz (11) compared white and Negro students from two Southern universities and found no significant differences regarding authoritarianism between the two ethnic groups. Further, he found only nine percent of the total sample were "true authoritarians," and also that intolerance of ambiguity was probably not highly related to authoritarianism. Christie and Garcia (2) compared students in California with a Southwestern city and found that students from the Southwestern city had higher F scale scores. Davids (4) studied 20 college students and found no significant relationship between F scale scores and tolerance of either ambiguous visual or auditory stimuli. Frenkel-Brunswik (6, 7) pointed out that prejudiced persons were less able to tolerate ambiguity. Maslow (14) discussed "self-actualizing people" and reported that such persons had clear perceptions of reality, could tolerate ambiguous situations, and were readily able to assert their independence should the situation so arise. Snygg and Combs (15) described an "adequate personality" in similar terms. Studies of visual perception of reality reported by Koffka (10) and Kelley (9) indicated that reality derived its characteristics from the total situation, including the perceiver, rather than from the stimuli alone. Finally, Stevens (16) said that the role of the listener needed to be studied by "varying the stimulus and/or by varying the listener."

Three major hypotheses were developed in order to aid in the exploration of the specific aspects of the problem. Each of these major hypotheses had several related sub-hypotheses. These are:

1. Persons from a Southern rural cultural situation will be more authoritarian than persons from a Northern urban cultural situation. That is, persons from a Southern rural cultural situation will:

a) score higher on the F scale
b) be less able to perceive reality
c) be less able to tolerate ambiguity
d) be less able to assert their independence
e) be less able to resist suggestion
f) have fewer perceptions

2. Girls will be more authoritarian than boys. That is, girls will:
   a) score higher on the F scale
   b) be less able to perceive reality
   c) be less able to tolerate ambiguity
   d) be less able to assert their independence
   e) be less able to resist suggestion
   f) have fewer perceptions

3. There will be a pattern of authoritarian behavioral characteristics which will indicate the existence of a syndrome. Specifically, there will be a positive relationship between the following factors:
   a) F scale scores and suggestibility
   b) F scale scores and fewer perceptions
   c) Suggestibility and fewer perceptions
   d) Perception of reality and assertion of independence
   e) Tolerance of ambiguity and assertion of independence
   f) Tolerance of ambiguity and perception of reality

Further, there will be a negative relationship between the following factors:
g) F scale scores and perception of reality
h) F scale scores and tolerance of ambiguity
i) F scale scores and assertion of independence
j) Suggestibility and perception of reality
k) Suggestibility and tolerance of ambiguity
l) Suggestibility and assertion of independence
m) Fewer perceptions and perception of reality
n) Fewer perceptions and tolerance of ambiguity
o) Fewer perceptions and assertion of independence

## The Procedure

The author developed a non-verbal sound test (8) after a series of seven pilot studies involving 486 persons, and then administered this test, together with tests of hearing acuity and the F scale, to groups of high-school students in rural Alabama and urban Michigan.

The sound test consisted of a series of questions and thirty-four non-verbal sounds recorded on magnetic tape. The questions were designed to procure specific information (age, sex, grade in school, etc.,) from each of the students participating in the study. Twenty sounds required a response to an aural stimulus in a situation unstructured except for time. Fourteen sounds, each of which was played two times, required responses in a more highly structured situation. Included also were two sample sounds designed to acquaint the students with the general procedure and style of the test. The tape was played at a constant speed and volume.

Each sound required approximately ten seconds of playing time and was followed by exactly eighteen seconds of silence. After hearing each of the first twenty sounds the subjects wrote out in a word or short phrase a description of the sound. Those sounds which were accurately described i.e., the description listed by the student was "correct," were considered as correct perceptions of reality. Those sounds which were not described, i.e., the student did not write out any kind of description in the space provided, were considered as "not perceived." In other words, for the purposes of this study unless a particular aural stimulus evoked a description of that sound, it was considered as not perceived.

The last fourteen sounds, each of which was played two times, required responses in a more structured situation. That is, each of these sounds had four alternative descriptions from which the student could choose, and one of these was suggested as being correct. Actually, none of the alternatives were correct descriptions of the particular sounds, although the last choice in each series was "something else," thus a subject could respond in that manner if he desired. If a subject, after hearing a sound, indicated his selection as that one which corresponded with the suggested choice, this was scored as indicating suggestibility. If a student decided that none of the descriptions listed were correct and selected "something else" as his choice, this was scored as an indication of his ability to assert his independence. After all fourteen sounds had been heard, they were replayed, and each subject was asked to indicate his degree of certainty about his initial perception, and this served as an indication of his ability to tolerate an ambiguous situation. That is, since none of the structured alternatives were correct except the "something else," and this was only partially correct since it did not describe the actual source of the sound, the degree of certainty a subject had about these ambiguous alternatives was ascertained.

For the purposes of this study, then, a perception of reality represented the correct description of an aural stimulus. Suggestibility represented an individual's response which corresponded to a suggested response for a particular aural stimulus. Tolerance of ambiguity represented the degree of certainty an individual had about his initial aural perception upon hearing certain sounds for a second time. Assertion of independence represented an individual's selection of the response "something else" when presented with a particular aural stimulus and three other alternative responses, none of which were correct, but one of which was suggested as being correct. Finally, those sounds which were not described were considered as not perceived.

The F (Fascist) scale (1), a commonly used device in psychological research, was selected as an instrument to determine an individual's tendency to be authoritarian. The instrument employed in this study consisted of 27 items administered as a "Public Opinion Questionnaire," the subjects indicating the extent of their agreement or disagreement with the various items. High F scale scores indicated a tendency toward authoritarianism.

The final group studied consisted of 22 girls and 42 boys from urban Michigan, and 48 girls and 43 boys from rural Alabama. The mean age of the Michigan sample was 16.31 years, and the mean age of the Alabama sample was 16.40 years. All subjects were Caucasian and had lived ten or more years in their particular locale. Only the results for those persons who passed a pure tone audiometer screening test were included in this report.

The subjects from Michigan came from a large metropolitan area in the southeastern section of the state. The school, located centrally in this automobile manufacturing city, had approximately 2000 students in grades nine through twelve. The student body was a heterogenous group representing different races, nationalities, and religious affiliations, as well as deviations in social and economic status. Two groups of ninth-grade and two groups of twelfth-grade students from this area were included in this study.

The subjects from Alabama came from a small, rural community situated about 55 miles southwest of Chattanooga, Tennessee, on top of what was commonly known as Sand Mountain. The population was exclusively white and the four churches in the immediate vicinity were all Protestant. The enrollment of the consolidated school from which this group came was approximately 750 students, grades one through twelve. Ninety-eight percent of the school population were bus-transported. The student body was essentially a homogeneous group, all white, and almost exclusively Protestant in their religious affiliation. Most of the students lived on small farms, thirty to forty acres in size. Two groups of ninth-grade and two groups of twelfth-grade students from this area were included in this study.

The statistical differences between the groups from these two cultural situations and between sexes was compared by computing t for the following scores: authoritarianism as depicted by the F scale, perception of reality, tolerance of ambiguity, assertion of independence, and number of sounds not perceived. Further, coefficients of correlation (r) between pairs of scores for students from both cultural situations, from both groups combined, for all of the boys, and for all of the girls combined were also computed.

The Results

Table I shows a comparison of mean scores of the Southern rural and Northern urban group performances on the F scale, perception of reality, tolerance of ambiguity, suggestibility, assertion of independence, and sounds not perceived, as ascertained in this study.

In general, persons from the Southern rural group scored markedly higher on the F scale (mean of 5.20) than did those persons from the Northern urban area (mean of 4.02). Further, those persons from the Southern rural group also had fewer perceptions of reality (mean of 5.74) than persons from the Northern urban group (mean of 8.47), were more suggestible (Southern rural group mean of 7.09, Northern urban group mean of 6.59), were less able to assert their independence (Southern rural group mean of .34, Northern urban group mean of 1.47), and had fewer perceptions, i.e., more sounds not perceived (Southern rural group mean of 4.44, Northern urban group mean of 2.25). Those persons from the Southern rural group, however, were more able to tolerate ambiguity (mean of 27.52) than were members of the Northern urban group (mean of 25.72), as measured in this study. All of these differences were statistically significant at the five percent level or higher except that one between mean scores on the suggestibility index.

Table II shows a comparison of the mean scores of boys from these two areas combined and of the girls on these same factors. These data indicate that there is a tendency among the girls tested in this particular study toward a pattern of authoritarian behavior characteristics, although none of the mean scores are significantly greater than those of the boys. Girls did score higher on the tolerance of ambiguity measure, but this difference was not statistically significant either.

Table III shows the coefficients of correlation between various pairs of the following scores: F scale, tolerance of ambiguity, perception of reality, assertion of independence, suggestibility, and number of sounds not perceived, as measured in this study. The correlation coefficients (r) were calculated for the Southern rural sample, the Northern urban sample, for both groups combined, for all of the boys, and for all of the girls. In this table those statistics which are underlined once (__) represent relationships in the predicted direction, whereas those which are underlined twice (__) represent relationships which are in the direction predicted and which are statistically significant at the five percent level of confidence.

There are several observations about Table III.

## TABLE I

## COMPARISON OF MEAN SCORES OF SOUTHERN RURAL AND NORTHERN URBAN GROUPS ON SIX BEHAVIORAL CHARACTERISTICS

| Group | F | R | Am | S | I | NP |
|---|---|---|---|---|---|---|
| Northern Urban | 4.02 | 8.47 | 25.72 | 6.59 | 1.47 | 2.25 |
| Southern Rural | 5.20 | 5.74 | 27.52 | 7.09 | .34 | 4.44 |
| t | 10.63* | 4.80* | 2.09* | 1.24 | 6.64* | 4.87* |

F = F scale scores
R = Perception of reality
Am = Tolerance of Ambiguity
S = Suggestibility
I = Assertion of independence
NP = Sounds not perceived
* = Differences significant at 5% level

## TABLE II

## COMPARISON OF MEAN SCORES OF BOYS AND GIRLS ON SIX BEHAVIORAL CHARACTERISTICS

| Group | F | R | Am | S | I | NP |
|---|---|---|---|---|---|---|
| Boys | 4.64 | 7.68 | 26.32 | 6.69 | .95 | 3.20 |
| Girls | 4.83 | 7.19 | 27.36 | 7.11 | .64 | 3.94 |
| t | 1.28 | 1.40 | 1.24 | 1.05 | 1.46 | 1.43 |

F = F scale scores
R = Perception of reality
Am = Tolerance of ambiguity
S = Suggestibility
I = Assertion of independence
NP = Sounds not perceived

## TABLE III

### COEFFICIENTS OF CORRELATION BETWEEN VARIOUS BEHAVIORAL CHARACTERISTICS FOR NORTHERN URBAN GROUP, SOUTHERN RURAL GROUP, BOTH GROUPS COMBINED, ALL BOYS, AND ALL GIRLS

| Two Traits Correlated | | No. Urban North | No. Rural South | No. Combined Groups | No. All Boys | No. All Girls |
|---|---|---|---|---|---|---|
| | | 64 | 91 | 155 | 85 | 70 |
| F Scale | Suggestibility | -.16 | .21 | .12 | .15 | .07 |
| F Scale | Not Perceived | -.09 | .00 | .21 | .15 | .24 |
| Suggestibility | Not Perceived | .13 | .29 | .26 | .28 | .28 |
| Reality | Independence | .00 | .10 | .19 | .24 | .10 |
| Ambiguity | Independence | -.04 | .26 | .00 | -.03 | .06 |
| Ambiguity | Reality | -.04 | .00 | -.08 | -.14 | .01 |
| F Scale | Reality | .08 | -.14 | -.28 | -.30 | -.24 |
| F Scale | Ambiguity | -.18 | -.20 | -.03 | -.02 | -.07 |
| F Scale | Independence | .16 | -.29 | -.30 | -.35 | -.21 |
| Suggestibility | Reality | .07 | -.07 | -.05 | .01 | -.10 |
| Suggestibility | Ambiguity | -.10 | -.16 | -.12 | -.09 | -.17 |
| Suggestibility | Independence | -.41 | -.22 | -.30 | -.42 | -.15 |
| Not Perceived | Reality | -.48 | -.48 | -.55 | -.49 | -.61 |
| Not Perceived | Ambiguity | .25* | .08 | .18* | .45* | -.13 |
| Not Perceived | Independence | .00 | .09 | -.18 | -.15 | -.19 |

— Indicates relationship in direction hypothesized.

═ Indicates relationship in direction hypothesized and significant at 5% level.

* Indicates relationship in direction opposite from that which was hypothesized and significant at 5% level.

The first, and perhaps the most important, is the fact that none of the correlation coefficients are very high, even though several are statistically significant.

Although these coefficients of correlation are not high, there seems to be some consistency or pattern of relationships which may be worthy of consideration. For instance, in the Southern rural group eleven of the fifteen relationships are in the direction predicted, and six of these are statistically significant. On the other hand, the Northern urban group has only three relationships which are statistically significant, and one of these is in the direction opposite from that which was hypothesized.

If both groups are combined and correlation coefficients calculated for the total sample of 155 persons, again the pattern of consistency seems to emerge. Those persons who tend toward authoritarianism as measured by the F scale seemed to be less able to perceive reality (r of -.28), less able to assert their independence (*r* of -.30), and had fewer perceptions (*r* of .21). Also, those persons who had fewer perceptions tended to be more suggestible (*r* of .26).

If one examines those coefficients of correlation between these various factors for all of the girls it seems that there is some very definite pattern apparent since all fifteen coefficients are in the direction predicted, although only four of these are statistically significant. The same pattern seems to prevail between pairs of scores for the boys, and although there are fewer relationships in the predicted direction, six of these coefficients of correlation are statistically significant.

It must be mentioned, however, that a major portion of the size of the r's between assertion of independence and suggestibility scores and between perception of reality and sounds not perceived scores is probably due to the fact that the design of the test was such that one of either of these pairs of scores could, under certain circumstances, cancel out the other. That is, if a person did not describe a sound, he could not have a correct perception of reality, thus this must account for at least some of the higher coefficients of correlation between perception of reality scores and sounds not perceived. Likewise, a person who selected the suggested response could not select "something else," and thus assert his independence, as measured in this study; therefore, these higher coefficients are probably due partly to the structure of the test itself.

## Discussion of the Results

Persons from the Southern rural cultural situation scored higher on the F scale, had fewer correct perceptions of reality, were less able to resist suggestion, and also had fewer perceptions. It seems that these data, then, are sufficient to accept, at least tentatively, the first hypothesis, namely, that persons from a Southern rural cultural situation will be more authoritarian than persons from a Northern urban cultural situation.

And although the girls scored higher on the authoritarian index, had fewer perceptions of reality, were less able to resist suggestion, were less able to assert their independence, and had fewer perceptions, none of these scores differed significantly from those of the boys. These data seem insufficient to accept the second hypothesis, therefore, that girls will be more authoritarian than boys, even though there does seem to be some evidence of a pattern of characteristics tending in such a direction.

Of the 75 coefficients of correlation calculated to portray the various relationships investigated in this particular study, 54 were in the direction predicted. Of this number, however, only 26 were significant statistically, and of these 26, nine were of a questionable nature (i. e., those between perception of reality and sounds not perceived and between suggestibility and assertion of independence). Only three coefficients of correlation in the direction opposite from that which was hypothesized were statistically significant, and all of these were between pairs of scores portraying the relationship of tolerance of ambiguity to number of sounds not perceived.

Although the data in this study are insufficient to accept the sub-hypotheses that tolerance of ambiguity is negatively related to certain authoritarian behavioral characteristics, something more significant may be inherent in these same data. Those persons who scored lower on the F scale, had more correct perceptions of reality, and who had more perceptions also tended to be *more* certain (i. e., *less* tolerant of ambiguity). However, these persons were also *more correct*; they were more effective in their perceptual processes. If they were more certain and less tolerant, they were so justifiably. This might be interpreted as a form of self-confidence or assurance which comes from having accurate perceptions and knowing they are accurate. This would hardly seem to be an undesirable characteristic, although it may be an authoritarian one. Also, it may be that this test did not actually measure tolerance of ambiguity as purported. The concept of tolerance of ambiguity apparently needs more study.

It would appear from these data, then, that there is some evidence supporting the hypothesis that there is a syndrome of authoritarian behavioral characteristics, although these data are more frequently not statistically significant than otherwise. There is an unmistakable pattern evident, but even though it is apparent, it is slight. Thus, the idea that there is an interrelationship of characteristics which would indicate the existence of a syndrome seems more plausible logically that tenable experimentally, at least as measured in this study. The characterization of the authoritarian personality as it has been discussed in other research seems to exist, but in a very indefinite and incomplete sense. It would seem that perhaps this theoretical characterization has been overdrawn, especially as regards intolerance of ambiguity.

## Conclusions

Persons from different cultural situations seem to perceive the same sounds differently.* Those persons from the Northern urban area who participated in the present study apparently had more perceptions and were more accurate in their perceptions.

Further, it seems that authoritarians are less "open" in the way they perceive aural stimuli. In other words, since they are less accurate, apparently have fewer perceptions, and are less able to assert their independence, among other things, those persons who are authoritarians do not seem to be as effective in the perceptive process as are nonauthoritarians.

Conversely, the nonauthoritarians seem to have more perceptions, to be more accurate in perceiving aural stimuli, and assert their independence more effectively than authoritarians. Because of these specific differences in perceptual abilities, it appears that nonauthoritarians are more "open" to aural stimuli than authoritarians.

## Suggested Research

During the course of the present study several related problems emerged. Some which the author feels deserve investigation are described:

1. How do persons from Northern rural and Southern urban cultural situations perceive aural stimuli?
2. What is the relationship of tolerance of ambiguity to authoritarianism?
3. Are the "authoritarian teacher" and the "authoritarian personality" the same?
4. Is there a difference between "direct" authoritarianism and "manipulation"?

## REFERENCES

1. Adorno, T. W., et al. The Authoritarian Personality (New York: Harper and Bros., 1950).
2. Christie, Richard and Garcia, J. "Subcultural Variation in Authoritarian Personality," Journal of Abnormal and Social Psychology, XLVI (October 1951), pp. 457-69.
3. Christie, Richard and Jahoda, Marie. Studies in the Scope and Method of "The Authoritarian Personality," (Glencoe, Ill.: The Free Press, 1954).
4. Davids, A. "Some Personality and Intellectual Correlates of Intolerance of Ambiguity," Journal of Abnormal and Social Psychology, LI (November 1955), pp. 415-20.
5. Drucker, Melvin B. Authority and Ingroup Standards, and Conformity Behavior, unpublished Ph. D. dissertation, George Peabody College for Teachers, 1956.
6. Frenkel-Brunswik, Else. "Tolerance Toward Ambiguity as a Personality Variable," American Psychologist, III (July 1948), pp. 268.
7. Frenkel-Brunswik, Else. "Intolerance of Ambiguity as an Emotional and Perceptual Variable," Journal of Personality, XVIII (September 1949), pp. 108-43.
8. Frymier, Jack R. The Relationship of Certain Behavioral Characteristics to Perception, unpublished Ed. D. dissertation, University of Florida, 1957.
9. Kelley, Earl C. Education for What is Real (New York: Harper and Bros., 1947).
10. Koffka, K. Principles of Gestalt Psychology (New York: Harcourt, Brace and Co., 1935).
11. Koontz, Miriam F. A Comparison of False Nonauthoritarians in Two Ethnic Groups, unpublished Ph. D. dissertation, George Peabody College for Teachers, 1955.
12. Masling, Joseph H. "How Neurotic is the Authoritarian?" Journal of Abnormal and Social Psychology, XLIX (April 1954), pp. 316-18.
13. Maslow, A. H. "The Authoritarian Character Structure," Journal of Social Psychology, XVIII (November 1943), pp. 401-11.
14. Maslow, A. H. Motivation and Personality (New York: Harper and Brothers, 1954).
15. Snygg, Donald and Combs, Arthur W. Individual Behavior (New York: Harper and Bros., 1949).
16. Stevens, S. S. (ed.) Handbook of Experimental Psychology (New York: John Wiley and Sons, 1951), p. 1068.
17. Titus, H. Edwin and Hollander, E. P. "The California F Scale in Psychological Research: 1950-1955," Psychological Bulleton, LIV (January 1957), pp. 47-64.

---

*Although this particular paper reports essentially a quantitative analysis, these same subjects' perceptions have also been analyzed qualitatively, i. e., according to kind.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 2, December 1959)

# GROUP STRUCTURE, ANXIETY, AND PROBLEM-SOLVING EFFICIENCY

FRANK W. BANGHART
University of Virginia

THE PURPOSE of this study was to compare the relative differences regarding the influence of anxiety on problem-solving, between cooperative and non-cooperative groups.

## Background

There is evidence in the literature to suggest that anxiety level does influence learning and performance. Castaneda and others (4) infer that "the tendency for the high anxious subjects to perform more poorly in comparison to low anxious subjects increases as the difficulty of the task increases." The difference found by Castaneda et al., was statistically significant at the .025 level. The difference was based upon the high anxious child doing more poorly on difficulty tasks than did the low anxious child. However, the high anxious child apparently did no better than the low anxious child on less difficult tasks. Palermo et al., (6) studied the relationship between anxiety and performance on a complex learning task and reported that the non-anxious subjects were supeior to the anxious subjects in every block of trials. The suggestion has been made in the Palermo et al., study that "the effect of increases in motivation on performance depends upon the relative strength of the correct and incorrect responses aroused by the experimental situation. Calvin et al., (3) report no relationship between anxiety and intelligence, but do report that anxiety does significantly inhibit performance. Sarasen (7) in a study of the effects of anxiety on two kinds of failure on social learning reported no significant difference between the control and experimental groups. Yet Flanders (5) concludes that "student behavior associated with interpersonal anxiety takes priority over behavior oriented toward the achievement problem." Ausubel et al., (1) reporting on a study dealing with anxiety and the learning process suggest that "the low anxiety group was significantly superior to the high anxiety group on the first trial of the maze, but this superiority was not maintained over the course of ten trials."

The above studies suggest that there are factors other than anxiety involved in situations where anxiety is a significant influence. In other words, anxiety per se is not necessarily the sole determining factor, but, rather, it becomes a significant influence under certain circumstances. It was the purpose of this study to investigate the relative influence of anxiety when groups are arranged cooperatively and non-cooperatively.

## Procedure

Twenty-four university students were given the Taylor Anxiety Scale. The subjects were then assigned to either a cooperative or non-cooperative experimental group. The basic difference between the two groups being in the free exchange of ideas and information on the part of the cooperative group. No deliberate communication was permitted for the non-cooperative group.

The experimental apparatus [described in detail elsewhere (2)] consisted of ten colored lights arranged on a panel. The problem involved the subjects' prediction of which light would come on next. The light series were predetermined and automatically controlled. The total experimental situation consisted of three problems; the first of which might be considered a "warm-up" task. Problem two consisted of a simple numerical solution (1, 3, 5, 7, 9); problem three required a place-color solution (red-red; green-green; etc.,) and involved a shift from a numerical to a color set. In terms of solution time and problem-solving efficiency, problem three was approximated fifty percent more difficult than problem two.

## Results

Table I summarizes the data in terms of problem-solving time and anxiety. Correlations were run between problem-solving time and anxiety for problems two and three for cooperative and non-cooperative groups. Table I suggests

---

*This study was made possible by a research grant from the Group Psychology Branch, Office of Naval Research, Contract No. NONR 474 (8).

### TABLE I

### CORRELATION BETWEEN PROBLEM-SOLVING TIME AND ANXIETY

| | | | |
|---|---|---|---|
| Non-Cooperative (P 2) | Problem-Solving Time | Anxiety | r = .06 |
| Cooperative (P 2) | Problem-Solving Time | Anxiety | r = .10 |
| | | p = .448 | |
| Non-Cooperative (P 3) | Problem-Solving Time | Anxiety | r = .04 |
| Cooperative (P 3) | Problem-Solving Time | Anxiety | r = .55 |
| | | p = .047 | |

P 2 = Problem two
P 3 = Problem three
p = probability of derived correlation coefficient

### TABLE II

### CORRELATION BETWEEN PROBLEM-SOLVING EFFICIENCY AND ANXIETY

| | | | |
|---|---|---|---|
| Non-Cooperative (P 2) | Problem-Solving Efficiency | Anxiety | r = -.35 |
| Cooperative (P 2) | Problem-Solving Efficiency | Anxiety | r = -.45 |
| | | p = .352 | |
| Non-Cooperative (P 3) | Problem-Solving Efficiency | Anxiety | r = .01 |
| Cooperative (P 3) | Problem-Solving Efficiency | Anxiety | r = -.48 |
| | | p = .068 | |

P 2 = Problem two
P 3 = Problem three
p = probability of derived correlation coefficient

that there was very little correlation (.06 and .10) between problem-solving time for either the non-cooperative or cooperative groups on problem two. Since problem two can be considered to be an "easy" problem, the results are consistent with the literature on the subject. However, our interest in this study was on the relative influence of anxiety between cooperative and non-cooperative groups. On the basis of this study no difference can be assumed (p = .448).

For problem three, which was fifty percent more difficult than problem two, an interesting discrepancy exists between the two groups (r = .04 for non-cooperative groups; r = .55 for cooperative groups, p = .047). One possible explanation of the discrepancy between the two groups on problem three might involve the thinking of the Flander's article (5). The suggestion that the non-cooperative groups were problem (achievement) centered, whereas the cooperative groups were interpersonal centered, the assumption is that anxiety interferes more when interpersonal relations are involved. This is speculation, of course, and needs to be empirically checked out.

Table II summarizes the data dealing with the relationship between problem-solving efficiency and anxiety. Efficiency in this case is defined as the ratio of correct to total responses.

Again, the correlations were run for problems two and three for cooperative and non-cooperative groups.

Reference to Table II suggests some interference (r = -.35 and -.45) by anxiety with problem-solving efficiency. However, the relative influence does not seem to differ significantly between the two groups (p = .352). (Reference might be made to the relative influence of anxiety on problem-solving efficiency and problem-solving time for problem two.)

For problem three a difference again emerges (p = .068) between the two groups. Attention is called to the decrease of anxiety influence, for the non-cooperative groups, from problem one (easy) to problem two (hard). Again, any attempt to explain the discrepancies between the two groups is in the realm of conjecture, yet attention should at least be called to other studies which might furnish a plausible explanation. A combination of the suggestions made by the Flanders (5) study and the Ausubel (1) study that the achievement centered subjects are less influenced by anxiety than are interpersonal relations centered subjects (Flanders), and that the discrepancy between high and low anxiety subjects does not hold up with time (Ausubel) might account for the findings of Tables I and II. Certainly some follow-up of these suggestions is indicated.

## Summary

An experimental study was made, using twenty-four university students, on the relative difference regarding the influence of anxiety on problem-solving between subjects assigned to cooperative and non-cooperative groups. Correlations were made between problem-solving time and anxiety and between problem-solving efficiency and anxiety. Tests were made regarding the significance of the differences between cooperative and non-cooperative groups.

The results suggest a minimal influence (r = .06 and .10) of anxiety on problem-solving time for the "easy" problem for both cooperative and non-cooperative groups (p = .448). However, anxiety seemed to have a more pronounced influence on the cooperative group than on the non-cooperative groups for the "hard" problem (r = .04 and .55; p = .047).

In terms of efficiency, anxiety seemed to have about the same influence for both groups (r = .35 and -.45; p = .352) on the "easy" problem.

However, on the "hard" problem, anxiety seemed to be more influential with the cooperative group than with the non-cooperative group in terms of influencing efficiency (r = .01 and -.48; p = .068).

At most, the results reported in this paper should be considered as exploratory and tentative. Certainly much additional work needs to be done in this area. A word of caution should be added also regarding the technique of measuring efficiency by a gross device such as right responses to total responses. Presumably, one could get a better measurement of anxiety, if a greater disparity existed between high and low anxiety scores. The present range (6 to 32, M = 13), with possibly too many scores around the mean, may not have been sensitive enough to discriminate between high and low anxiety.

Again, it should be emphasized that more work needs to be done before anything beyond suggestion can be inferred.

## REFERENCES

1. Ausubel, D. P., and others. "Qualitative Characteristics in the Learning Process Associated with Anxiety," Journal of Abnormal and Social Psychology," XLVIII (October 1948).

2. Banghart, F. W., and others. "A Study of 'Interfering Variables' Upon Communication in Elementary School Children," Educational Research Bulletin, I (June 1957).

3. Calvin, A. D., and others. "Further Investigation of the Relationship Between Anxiety and Classroom Examination Performance," Journal of Educational Psychology, XLVIII (April 1957), pp. 240-4.

4. Castaneda, A., and others. "Complex Learning and Performance as a Function of Anxiety in Children and Task Difficulty," Child Development, XXVII (September 1956), pp. 327-

5. Flanders, N. A. "Personal Social Anxiety as a Factor in Experimental Learning Situations," Journal of Educational Research, XLV (October 1951), pp. 100-10.

6. Palermo, D. S., and others. "Relationship of Anxiety in Children to Performance in a Complex Learning Task," Child Development, XXVII (September 1956), pp. 333-7.

7. Saracen, I. G. "Effect of Anxiety and Kinds of Failure in Social Learning," Journal of Personnel, XXV (March 1957), pp. 383.

# The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1960 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# MULTIPLE FACTOR ANALYSIS OF ACADEMIC ACHIEVEMENT: A COMPARATIVE STUDY OF FULL-BLOODED INDIAN AND WHITE CHILDREN*

OMER JOHN RUPIPER**
University of Kansas

## I. Introduction

THIS STUDY had its origin in the studies conducted by the Bureau of Indian Affairs and the University of Kansas during the period 1951-55 (3, 13). This Indian Education Research Project is undoubtedly the most comprehensive study of Indian and white children which has been made. The program explored the educational status of Indian children and emphasized two functions: predictive testing and achievement testing. The purpose of the predictive testing was an attempt to investigate high-school graduates' academic skills and to predict their probable success if they consider further formal education. The achievement testing was designed to serve two purposes: (a) a continuing evaluation of the status of educational achievement of Indian children for administrative use, that is, the recording and interpreting of test scores for describing a pupil's educational achievement as measured by tests, and (b) the use of data obtained as a functional part of the instructional program. These studies showed that Indian children differed from white children as measured by achievement test results and that the smaller the proportion of full-blooded and non-English speaking children in a specific group of Indian children the higher the group achieved. It was recognized that these differences in educational achievement depended on differences in experiential, environmental, and cultural factors and not necessarily on innate ability alone.

The present study was an attempt to isolate and compare the factor structures and sets of factor patterns obtained from the achievement test scores for full-blooded Indian children and white children. The purpose of this study was to investigate the degree of relationship between the subtests of the measuring instrument, to determine the number of independent factors which account for the scores on the subtests, how heavily each of the subtests is weighted with the factors, and which factors appear as a whole for the different groups of subjects. A comparison of factors by sex, grade level, and race was made.

Although there has been a great deal of research in the literature which concerns intelligence and educational achievement of American Indians, no studies were evident in which factor analysis had been applied to achievement test scores for the purpose of comparing full-blooded Indian and white children on large samples. Perhaps the only attempt was made by Harris (33) in which correlations were obtained between 15 language skill variables in four groups of fifth-grade Indian students. The intercorrelations were too irregular for any definite factors to emerge. Experimental studies on intelligence with Indians (20, 26, 27, 28, 29, 30, 34, 36, 39, 46) indicate that differences in test performance between Indian and white children are not necessarily due to differences in innate capacity. The evidence indicates that white children and mixed-blooded Indian children did better where language was vital. The results showed that Indians scored below age and that females performed better than males on the intelligence tests used, but these results seem meager and inconclusive.

Anthropological studies (37, 39, 40, 41) indicated that there were no significant differences in results of performance intelligence tests for Indian children as compared to white children. Intelligence tests of a verbal nature, requiring writing, resulted in lower IQ's for Indian children. The educational achievement of Indians in public schools, federal schools, and mission schools as measured by test scores resulted in a hierarchical order, respectively (2, 13, 42, 43, 52). Indian children with highest degree of Indian blood and non-English speaking ability at the time of entering school obtained the lowest test results. It was believed that cultural factors played an important role in the learning process of these children.

Factor analytic studies on intelligence and age with white children (1, 4, 5, 7, 9, 14, 17, 19, 23, 24, 25, 38, 44, 45, 47, 48) emphasized that a general factor was established early in life. However, conflicting evidence existed as to the first factor changing with increasing age. Where factor patterns did change, there was no consistent trend in evidence. Factor studies regarding levels and sex for white children (35, 49, 50, 51) revealed a general agreement as to no differentiation in factor patterns between boys and girls. There was indication of conflicting evidence regarding the theory of the general factor playing a less important role as special abilities develop. The factor patterns change relatively little from grade level to grade level, and where changes did occur the patterns did not exhibit any systematic trend. Mental factors seemed to depend on background, education, and experience of the subjects involved.

---

*All footnotes will be found at end of article.

Where educational skills for white children were factor analyzed (6, 8, 10, 11, 12, 15, 16, 18, 32, 53) it was apparent that educational skills were rather complex with regard to the identification of component factors, that boys and girls vary according to different academic skills, and the more simple and relevant the test the more clearly it produced the factor. Coleman (11) pointed out that a pupil's basic learning difficulty can be diagnosed at an earlier age than was formerly believed and that school achievement depended upon motivation and other factors.

In consideration of the review of literature, five important conclusions were drawn from the results: 1) Factors other than innate capacity were operating on test performance of both Indian and white children; 2) because of contradictory findings, no conclusive statements can be made as to sex and racial superiority on test performance; 3) factor patterns change relatively little from grade to grade with no significant changes between sex; 4) a common factor is established relatively early in life with doubt as to its role becoming minimized as specific abilities develop with increasing age, and 5) because of the complexity of educational skills, it was difficult to identify their component factors.

### Hypotheses

On the basis of these studies, one hypothesis concerning differences in achievement raw score means by grade level as to race and sex was developed. Five hypotheses were formulated pertaining to the number, size, and pattern of factors of achievement scores as measured by the California Achievement Test Batteries. The hypotheses are:

1. No differences exist in achievement raw score means on the subtests for the elementary, intermediate, and advanced grade levels for each race by sex.
2. The same number of factors will appear at each grade level for the race groups by sex.
3. The factor loadings will be of the same magnitude for all groups.
4. Higher factor loadings of a verbal nature will appear for white children than for full-blooded Indian children.
5. Higher factor loadings of a numerical nature will appear for white children than for full-blooded Indian children.
6. A common factor will appear throughout all the subtests for all groups.

## II. Procedure

### The Testing Program

A fall achievement testing program was conducted by the United States Bureau of Indian Affairs and the University of Kansas during the period 1951-55. The California Achievement Test Battery was used throughout the entire testing program for grades four through twelve. The children attending federal schools, public schools, and mission schools were tested in the Albuquerque and Phoenix areas in 1951, Aberdeen area in 1952, Billings area in 1953, and the Anadarko and Muskogee areas in 1954. The tests were administered by Indian Bureau and public and mission school personnel in their respective schools. The public and mission schools included were those which operated in the vicinity of the federal schools included in this study and they enrolled a considerable proportion of Indian pupils. All answer sheets were sent to the University of Kansas for inspection, necessary processing, and scoring. The raw scores of each subtest for each individual were recorded and tabulated. These were punched on IBM cards with background information for each child.

### The Population

The geographical locations of the administrative areas of the United States Bureau of Indian Affairs in which testing was conducted were: 1) Phoenix area—Arizona, exclusive of the Navajo and Hopi, 2) Albuquerque area—New Mexico and southern Colorado, exclusive of the Navajo, 3) Aberdeen area—South Dakota, North Dakota, and northeastern Nebraska, 4) Billings area—Montana and Wyoming, 5) Muskogee area—eastern Oklahoma and Mississippi, and 6) Anadarko area—western Oklahoma and Kansas. Between the years 1951 and 1954, inclusive, 13,636 Indian children and 9922 white children were tested in federal, public, and mission schools. These 23,608 children lived in rural areas either on farms, ranches, reservations, or small communities.

The aggregate of full-blooded Indian children attending federal schools and white children attending public schools combined in all areas were chosen for this study. The total number of subjects, 14,888, were separated as to full-blooded Indian children and white children by sex and

## TABLE I

## DISTRIBUTION OF SUBJECTS BY RACE, SEX, AND GRADE LEVELS

| | Full-Blood Indian Children | | | |
|---|---|---|---|---|
| | Males | | Females | |
| | N | % | N | % |
| Elementary Level Grades 4, 5, 6 | 1223 | 21.32 | 1161 | 20.24 |
| Intermediate Level Grades 7, 8, 9 | 972 | 17.81 | 971 | 17.79 |
| Advanced Level Grades 10, 11, 12 | 551 | 14.91 | 624 | 16.89 |

| | White Children | | | |
|---|---|---|---|---|
| | Males | | Females | |
| | N | % | N | % |
| Elementary Level Grades 4, 5, 6 | 1728 | 30.13 | 1624 | 28.31 |
| Intermediate Level Grades 7, 8, 9 | 1817 | 33.30 | 1697 | 31.10 |
| Advanced Level Grades 10, 11, 12 | 1201 | 32.50 | 1319 | 35.70 |
| Total = 14,888 | | | | |

grade levels. To conform to the suggested grade placement for the achievement tests, the grade levels were combined into elementary (grades 4, 5, and 6), intermediate (grades 7, 8, and 9), and advanced (grades 10, 11, and 12). Table I shows the breakdown of the subjects and the number in each group with percentages. The approximate age ranges for the federal Indian children were from eight to 17 years for the elementary level, 10 to 19 years for the intermediate level, and 15 to 24 years for the advanced level. For public school white children the approximate age ranges were from eight to sixteen years for the elementary level, 11 to 19 years for the intermediate level, and 15 to 20 years for the advanced level.

## The Test Battery

The California Achievement Test Battery, Form AA, was utilized to measure the subjects' achievement in fundamental reading, arithmetic, and language skills through the entire testing program. These tests were chosen because they were already being used widely in Indian Bureau schools, they were selected by Indian Bureau personnel as being as close to content of the curriculum as any of the other achievement batteries, and they were machine scored. The tests of basic skills used were: a) reading vocabulary, b) reading comprehension, c) arithmetic reasoning, d) arithmetic fundamentals, e) mechanics of English and grammar, and f) spelling. Since the sections of each test contained such a small number of items, only total scores for each test were obtained rather than scores for each section of the test. The scoring formula was the number of right responses for each of the six tests. Since the California Achievement Test Battery is widely known and is used as a test of academic achievement, a discussion of its construction, reliability, and validity is unnecessary.

## The Factorial Analyses

The raw scores for each individual were punched on IBM cards. Tabulating machine operations were used in obtaining frequency distributions, sums of squares, and sums of cross products for each test by race, sex, and grade level groups.1 The means, medians, standard deviations, ranges, and indices of skewness were computed from the frequency distributions. There were only slight discrepancies with no differences between the median and mean greater than three raw score points. In many cases the differences were less than one. Many of the distributions were slightly skewed which indicated that the tests could have been somewhat too difficult for the students tested. However, the degree of skewness was not sufficiently large where the assumption of normality could not be met. A rather consistent trend was noted in the variability among races, sexes, and grade level groups. In general, the standard deviation is considerably larger on the Reading, Vocabulary, and Arithmetic fundamentals test than on the other four tests in the battery for the various groups. Pearson product-moment correlation coefficients were determined among the six tests by race, sex, and grade level. Twelve correlation matrices for the various groups were formulated, i.e., a separate correlation matrix was constructed for males and females at the elementary, intermediate, and advanced grade level groups for both full-blooded Indian children and white children. The correlations were all positive with the lowest correlation of r = .35 between the Reading vocabulary test and the Arithmetic fundamentals test at the advanced level for full-blooded Indian children. For the sample size N = 551 this correlation differs from zero at the $.\overline{01}$ level of confidence, as do all other coefficients of correlation in this study.

The correlation matrices were factored by the centroid method of analysis (22, 31, 54). The first extraction of factors was accomplished by using the highest correlation in each matrix as the estimated communality. After each factoring, adjustments were made in the diagonal cells and the process was continued until three factors were extracted. In order to obtain a better estimate of the communalities, the calculated communalities were used in the diagonal cells and the factoring was repeated. The new centroid factor loadings and the revised estimate of the communalities are listed in Tables IV through VII. After the second extraction of factors the residuals of the third factor were sufficiently small to cease factoring.

An oblique simple structure was obtained for the three factors. The method of extended vectors was utilized because, according to Thurstone (54), where three dimensions are used the method gives the solution in one diagram if a simple-structure solution exists. This method seemed most appropriate since a central reference vector existed on which all the test vectors have positive projections. The Extended Factor Matrices $E_0$ are shown in Table VIII through XIX. Upon oblique rotation of the centroid factor loadings for full-blooded Indian male children at all grade level groups, the third extracted factor rotated with the second did not reveal clear configurations, therefore, meaningful traces could not be drawn which would aid in the establishment and interpretation of these factors. The second factor residuals, range -.05 to .04, for full-blooded Indian male children at all levels are sufficiently small to yield a meaningful third factor. Apparently, the third factor loadings for this group are error residuals. As a result, only factors I and II were used in oblique rotation for these three grade levels in which a single plane, two-dimensional rotation was made. Tables VIII, XIX,

## TABLE II

## INTERCORRELATIONS BETWEEN THE RAW SCORES OF EACH TEST FOR FULL-BLOOD INDIAN CHILDREN BY SEX ON THE CALIFORNIA ACHIEVEMENT BATTERY

| | | Full-Blood Indian Children | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Males (N=1223) | | | | | | Females (N=1161) | | | | | |
| Level | Test[a] | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| Elementary | 1 | | 72 | 72 | 56 | 66 | 64 | | 76 | 69 | 54 | 65 | 65 |
| | 2 | 72 | | 71 | 56 | 69 | 63 | 76 | | 70 | 60 | 70 | 66 |
| | 3 | 72 | 71 | | 73 | 68 | 65 | 69 | 70 | | 71 | 65 | 65 |
| | 4 | 56 | 56 | 73 | | 54 | 48 | 54 | 60 | 71 | | 52 | 53 |
| | 5 | 66 | 69 | 68 | 54 | | 59 | 65 | 70 | 65 | 52 | | 64 |
| | 6 | 64 | 63 | 65 | 48 | 59 | | 65 | 66 | 65 | 53 | 64 | |
| | | (N=972) | | | | | | (N=971) | | | | | |
| Intermediate | 1 | | 68 | 61 | 51 | 64 | 71 | | 62 | 54 | 47 | 60 | 63 |
| | 2 | 68 | | 71 | 63 | 67 | 65 | 62 | | 63 | 59 | 64 | 59 |
| | 3 | 61 | 71 | | 78 | 70 | 56 | 54 | 63 | | 71 | 66 | 54 |
| | 4 | 51 | 63 | 78 | | 65 | 50 | 47 | 59 | 71 | | 60 | 48 |
| | 5 | 64 | 67 | 70 | 65 | | 62 | 60 | 64 | 66 | 60 | | 63 |
| | 6 | 71 | 65 | 56 | 50 | 62 | | 63 | 59 | 54 | 48 | 63 | |
| | | (N=551) | | | | | | (N=624) | | | | | |
| Advanced | 1 | | 68 | 53 | 35 | 55 | 63 | | 58 | 49 | 37 | 52 | 51 |
| | 2 | 68 | | 61 | 52 | 60 | 60 | 58 | | 61 | 49 | 61 | 57 |
| | 3 | 53 | 61 | | 70 | 63 | 52 | 49 | 61 | | 72 | 59 | 48 |
| | 4 | 35 | 52 | 70 | | 58 | 45 | 37 | 49 | 72 | | 53 | 46 |
| | 5 | 55 | 60 | 63 | 58 | | 55 | 52 | 61 | 59 | 53 | | 59 |
| | 6 | 63 | 60 | 52 | 45 | 55 | | 51 | 57 | 48 | 46 | 59 | |

Note.--Decimals omitted for convenience.

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE III

## INTERCORRELATIONS BETWEEN THE RAW SCORES OF EACH TEST FOR WHITE CHILDREN BY SEX ON THE CALIFORNIA ACHIEVEMENT BATTERY

| Level | Test[a] | White Children | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Males | | | | | | Females | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| | | (N=1728) | | | | | | (N=1624) | | | | | |
| Elementary | 1 | | 76 | 75 | 55 | 67 | 64 | | 77 | 75 | 59 | 67 | 67 |
| | 2 | 76 | | 73 | 57 | 67 | 65 | 77 | | 76 | 63 | 66 | 65 |
| | 3 | 75 | 73 | | 73 | 68 | 62 | 75 | 76 | | 75 | 65 | 63 |
| | 4 | 55 | 57 | 73 | | 56 | 51 | 59 | 63 | 75 | | 54 | 53 |
| | 5 | 67 | 67 | 68 | 56 | | 61 | 67 | 66 | 65 | 54 | | 61 |
| | 6 | 64 | 65 | 62 | 51 | 61 | | 67 | 65 | 63 | 53 | 61 | |
| | | (N=1817) | | | | | | (N=1697) | | | | | |
| Intermediate | 1 | | 76 | 67 | 54 | 65 | 67 | | 74 | 66 | 59 | 67 | 62 |
| | 2 | 76 | | 73 | 63 | 68 | 61 | 74 | | 71 | 65 | 68 | 56 |
| | 3 | 67 | 73 | | 75 | 68 | 57 | 66 | 71 | | 77 | 68 | 54 |
| | 4 | 54 | 63 | 75 | | 64 | 53 | 59 | 65 | 77 | | 67 | 52 |
| | 5 | 65 | 68 | 68 | 64 | | 61 | 67 | 68 | 68 | 67 | | 58 |
| | 6 | 67 | 61 | 57 | 53 | 61 | | 62 | 56 | 54 | 52 | 58 | |
| | | (N=1201) | | | | | | (N=1319) | | | | | |
| Advanced | 1 | | 72 | 63 | 49 | 58 | 61 | | 71 | 59 | 46 | 61 | 62 |
| | 2 | 72 | | 63 | 54 | 58 | 53 | 71 | | 62 | 53 | 61 | 56 |
| | 3 | 63 | 63 | | 75 | 67 | 53 | 59 | 62 | | 76 | 63 | 48 |
| | 4 | 49 | 54 | 75 | | 58 | 43 | 46 | 53 | 76 | | 59 | 46 |
| | 5 | 58 | 58 | 67 | 58 | | 60 | 61 | 61 | 63 | 59 | | 64 |
| | 6 | 61 | 53 | 53 | 43 | 60 | | 62 | 56 | 48 | 46 | 64 | |

Note.--Decimals omitted for convenience.

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE IV

## CENTROID FACTOR MATRICES AND OBTAINED COMMUNALITIES WITH PROPORTIONS OF VARIANCES CONTRIBUTED BY THE CENTROID FACTORS ($F_0$) (MALE)

| | | Full-Blood Indian Children | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Male | | | | | | |
| | | Factor Loadings | | | Factor Variances | | | $h^2$ |
| Level | Test[a] | I | II | III | I' | II' | III' | |
| Elementary | 1 | 832 | 123 | 036 | 692 | 015 | 001 | 708 |
| | 2 | 840 | 162 | 155 | 706 | 026 | 024 | 756 |
| | 3 | 888 | -192 | -131 | 789 | 037 | 017 | 843 |
| | 4 | 737 | -438 | 131 | 543 | 192 | 017 | 752 |
| | 5 | 788 | 100 | 096 | 621 | 010 | 009 | 640 |
| | 6 | 747 | 285 | -287 | 558 | 081 | 082 | 721 |
| Intermediate | 1 | 798 | 324 | -043 | 637 | 105 | 002 | 744 |
| | 2 | 833 | 079 | 149 | 694 | 006 | 022 | 722 |
| | 3 | 861 | -281 | 064 | 741 | 079 | 004 | 824 |
| | 4 | 787 | -353 | 043 | 619 | 124 | 002 | 745 |
| | 5 | 814 | -036 | -064 | 663 | 001 | 004 | 668 |
| | 6 | 769 | 317 | -107 | 591 | 100 | 011 | 702 |
| Advanced | 1 | 758 | 425 | 151 | 575 | 181 | 023 | 779 |
| | 2 | 805 | 184 | -060 | 648 | 034 | 004 | 686 |
| | 3 | 821 | -272 | 151 | 674 | 074 | 023 | 771 |
| | 4 | 723 | -425 | -166 | 523 | 181 | 028 | 732 |
| | 5 | 766 | -070 | -060 | 587 | 005 | 004 | 596 |
| | 6 | 723 | 203 | -075 | 523 | 041 | 006 | 570 |

Note.--Decimals omitted for convenience.

[a]Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE V

## CENTROID FACTOR MATRICES AND OBTAINED COMMUNALITIES WITH PROPORTIONS OF VARIANCES CONTRIBUTED BY THE CENTROID FACTORS ($F_0$) (FEMALE)

| | | Full-Blood Indian Children | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Female | | | | | | |
| | | Factor Loadings | | | Factor Variances | | | |
| Level | Test[a] | I | II | III | I' | II' | III' | $h^2$ |
| Elementary | 1 | 830 | 175 | 104 | 689 | 031 | 011 | 731 |
| | 2 | 863 | 140 | -070 | 745 | 020 | 005 | 770 |
| | 3 | 859 | -236 | 104 | 738 | 056 | 011 | 805 |
| | 4 | 735 | -315 | 017 | 540 | 099 | 000 | 639 |
| | 5 | 793 | 166 | -122 | 629 | 028 | 015 | 672 |
| | 6 | 776 | 114 | 157 | 602 | 013 | 025 | 640 |
| | 1 | 750 | 284 | 078 | 563 | 081 | 006 | 650 |
| | 2 | 795 | 041 | 098 | 632 | 002 | 010 | 644 |
| | 3 | 816 | -276 | -039 | 666 | 076 | 002 | 744 |
| | 4 | 756 | -333 | 020 | 572 | 111 | 000 | 683 |
| | 5 | 814 | -032 | -157 | 663 | 001 | 025 | 689 |
| | 6 | 752 | 268 | -118 | 566 | 072 | 014 | 652 |
| | 1 | 675 | 258 | -167 | 456 | 067 | 028 | 551 |
| | 2 | 782 | 176 | -143 | 612 | 031 | 020 | 663 |
| | 3 | 820 | -326 | -084 | 672 | 106 | 007 | 785 |
| | 4 | 729 | -360 | 143 | 531 | 130 | 020 | 681 |
| | 5 | 769 | 109 | 108 | 591 | 012 | 012 | 615 |
| | 6 | 715 | 244 | 191 | 511 | 060 | 036 | 607 |

Note.--Decimals omitted for convenience.

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language 6. Spelling

TABLE VI

CENTROID FACTOR MATRICES AND OBTAINED COMMUNALITIES WITH PROPORTIONS OF VARIANCES CONTRIBUTED BY THE CENTROID FACTORS ($F_0$) (MALE)

| | | White Children | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Male | | | | | | |
| | | Factor Loadings | | | Factor Variances | | | $h^2$ |
| Level | Test[a] | I | II | III | I' | II' | III' | |
| Elementary | 1 | 859 | 176 | 170 | 738 | 031 | 029 | 798 |
| | 2 | 850 | 166 | 099 | 723 | 028 | 010 | 761 |
| | 3 | 887 | -213 | 141 | 787 | 045 | 020 | 852 |
| | 4 | 739 | -324 | -085 | 546 | 105 | 007 | 658 |
| | 5 | 787 | 065 | -099 | 619 | 004 | 010 | 633 |
| | 6 | 748 | 139 | -113 | 560 | 019 | 013 | 592 |
| Intermediate | 1 | 837 | 334 | 093 | 701 | 111 | 009 | 821 |
| | 2 | 864 | 109 | 172 | 746 | 012 | 030 | 788 |
| | 3 | 856 | -226 | 159 | 733 | 051 | 025 | 809 |
| | 4 | 784 | -360 | -026 | 615 | 130 | 001 | 746 |
| | 5 | 804 | -025 | -106 | 646 | 001 | 011 | 658 |
| | 6 | 739 | 142 | -199 | 546 | 020 | 040 | 606 |
| Advanced | 1 | 806 | 341 | 207 | 650 | 116 | 043 | 809 |
| | 2 | 787 | 203 | 197 | 619 | 041 | 039 | 699 |
| | 3 | 855 | -283 | 128 | 731 | 080 | 016 | 827 |
| | 4 | 750 | -334 | 089 | 563 | 112 | 008 | 683 |
| | 5 | 780 | 051 | -197 | 608 | 003 | 039 | 650 |
| | 6 | 699 | 167 | -197 | 489 | 028 | 039 | 556 |

Note.--Decimals omitted for convenience.

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling.

## TABLE VII

## CENTROID FACTOR MATRICES AND OBTAINED COMMUNALITIES WITH PROPORTIONS OF VARIANCES CONTRIBUTED BY THE CENTROID FACTORS ($F_0$) (FEMALE)

| | | White Children | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Female | | | | | | |
| | | Factor Loadings | | | Factor Variances | | | |
| Level | Test[a] | I | II | III | I' | II' | III' | $h^2$ |
| Elementary | 1 | 864 | 181 | -101 | 746 | 033 | 010 | 789 |
| | 2 | 866 | 078 | -068 | 750 | 006 | 005 | 761 |
| | 3 | 891 | -207 | 051 | 794 | 043 | 003 | 840 |
| | 4 | 764 | -294 | 135 | 584 | 086 | 018 | 688 |
| | 5 | 766 | 233 | 220 | 587 | 054 | 084 | 725 |
| | 6 | 756 | 164 | -017 | 572 | 027 | 000 | 599 |
| Intermediate | 1 | 834 | 274 | 122 | 696 | 075 | 015 | 786 |
| | 2 | 844 | 091 | 191 | 712 | 008 | 036 | 756 |
| | 3 | 854 | -246 | 109 | 729 | 061 | 012 | 802 |
| | 4 | 815 | -301 | -054 | 664 | 091 | 003 | 758 |
| | 5 | 811 | 009 | -068 | 658 | 000 | 005 | 663 |
| | 6 | 698 | 173 | -191 | 487 | 030 | 036 | 553 |
| Advanced | 1 | 796 | 291 | -226 | 634 | 045 | 051 | 730 |
| | 2 | 798 | 132 | -129 | 637 | 017 | 017 | 671 |
| | 3 | 824 | -357 | -097 | 679 | 127 | 009 | 815 |
| | 4 | 754 | -390 | 129 | 567 | 152 | 017 | 736 |
| | 5 | 796 | 072 | 194 | 634 | 005 | 038 | 677 |
| | 6 | 728 | 271 | 151 | 530 | 073 | 023 | 626 |

Note.--Decimals omitted for convenience.

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

# TABLE VIII

## MATRICES OF OBLIQUE ROTATION OF AXES FOR FULL-BLOOD INDIAN CHILDREN, MALES, ELEMENTARY GRADE LEVEL GROUP

Unrotated Factor Matrix $F_o$

| Test[a] | I | II |
|---|---|---|
| 1 | .832 | .123 |
| 2 | .840 | .162 |
| 3 | .888 | -.192 |
| 4 | .737 | -.438 |
| 5 | .788 | .100 |
| 6 | .747 | .285 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' |
|---|---|---|
| I | .40 | -.21 |
| II | 1.00 | 1.00 |

Matrix C = $\Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda pq$

| Factor | A | B |
|---|---|---|
| A | 1.000 | |
| B | .833 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B |
|---|---|---|
| 1 | .424 | -.051 |
| 2 | .463 | -.014 |
| 3 | .152 | -.371 |
| 4 | -.133 | -.581 |
| 5 | .386 | -.064 |
| 6 | .543 | -.125 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda p$

| Factor | A | B |
|---|---|---|
| I | .372 | -.206 |
| II | .929 | .979 |

Extended Factor Matrix $E_o$

| Test[a] | De | I | II | III |
|---|---|---|---|---|
| 1 | 1.20192 | 1.000 | .148 | .043 |
| 2 | 1.19048 | 1.000 | .193 | .185 |
| 3 | 1.12613 | 1.000 | -.216 | -.148 |
| 4 | 1.35685 | 1.000 | -.594 | .178 |
| 5 | 1.26904 | 1.000 | .127 | .122 |
| 6 | 1.33869 | 1.000 | .382 | -.384 |

[a] Test:--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE IX

## MATRICES OF OBLIQUE ROTATION OF AXES FOR FULL-BLOOD INDIAN CHILDREN, MALES, INTERMEDIATE GRADE LEVEL GROUP

Unrotated Factor Matrix $F_o$

| Test[a] | I | II |
|---|---|---|
| 1 | .798 | .324 |
| 2 | .833 | .079 |
| 3 | .861 | -.281 |
| 4 | .787 | -.353 |
| 5 | .814 | -.036 |
| 6 | .769 | .317 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' |
|---|---|---|
| I | .27 | -.30 |
| II | 1.00 | 1.00 |

Matrix C = $\Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda_{pq}$

| Factor | A | B |
|---|---|---|
| A | 1.000 | |
| B | .851 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B |
|---|---|---|
| 1 | .521 | .081 |
| 2 | .294 | -.163 |
| 3 | -.047 | -.516 |
| 4 | -.136 | -.564 |
| 5 | .178 | -.268 |
| 6 | .507 | .083 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda_p$

| Factor | A | B |
|---|---|---|
| I | .261 | -.287 |
| II | .966 | .958 |

Extended Factor Matrix $E_o$

| Test[a] | De | I | II | III |
|---|---|---|---|---|
| 1 | 1.25313 | 1.000 | .406 | -.054 |
| 2 | 1.20048 | 1.000 | .095 | .179 |
| 3 | 1.16144 | 1.000 | -.326 | .074 |
| 4 | 1.27065 | 1.000 | -.449 | .055 |
| 5 | 1.22850 | 1.000 | -.044 | -.079 |
| 6 | 1.30039 | 1.000 | .412 | -.139 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE X

## MATRICES OF OBLIQUE ROTATION OF AXES FOR FULL-BLOOD INDIAN CHILDREN, MALES, ADVANCED GRADE LEVEL GROUP

Unrotated Factor Matrix $F_o$

| Test[a] | I | II |
|---|---|---|
| 1 | .758 | .425 |
| 2 | .805 | .184 |
| 3 | .821 | -.272 |
| 4 | .723 | -.425 |
| 5 | .766 | -.070 |
| 6 | .723 | .203 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B |
|---|---|---|
| 1 | .647 | .143 |
| 2 | .433 | -.010 |
| 3 | .007 | -.534 |
| 4 | -.170 | -.645 |
| 5 | .180 | -.326 |
| 6 | .041 | -.054 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' |
|---|---|---|
| I | .34 | -.36 |
| II | 1.00 | 1.00 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda p$

| Factor | A | B |
|---|---|---|
| I | .322 | -.339 |
| II | .947 | .941 |

Matrix $C = \Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda pq$

| | A | B |
|---|---|---|
| A | 1.000 | |
| B | .785 | 1.000 |

Extended Factor Matrix $E_o$

| Test[a] | De | I | II | III |
|---|---|---|---|---|
| 1 | 1.31926 | 1.000 | .561 | .199 |
| 2 | 1.24224 | 1.000 | .229 | -.075 |
| 3 | 1.21803 | 1.000 | -.331 | .184 |
| 4 | 1.38313 | 1.000 | -.588 | -.230 |
| 5 | 1.30548 | 1.000 | -.091 | -.078 |
| 6 | 1.38313 | 1.000 | .281 | -.104 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals. 5. Language, 6. Spelling

## TABLE XI

## MATRICES OF OBLIQUE ROTATION OF AXES FOR FULL-BLOOD INDIAN CHILDREN, FEMALES, ELEMENTARY GRADE LEVEL GROUP

Extended Factor Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.20482 | 1.000 | .211 | .125 |
| 2 | 1.15875 | 1.000 | .162 | -.081 |
| 3 | 1.16414 | 1.000 | -.275 | .121 |
| 4 | 1.36054 | 1.000 | -.429 | .023 |
| 5 | 1.26103 | 1.000 | .209 | -.154 |
| 6 | 1.28866 | 1.000 | .147 | .202 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .04 | .19 | .14 |
| II | .36 | -1.00 | .18 |
| III | 1.00 | - .02 | -1.00 |

Matrix C = $\Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda_{pq}$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | -.345 | -.852 |
| B | -.345 | 1.000 | -.128 |
| C | -.852 | -.128 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | .189 | -.019 | .043 |
| 2 | .014 | .025 | .211 |
| 3 | .050 | .390 | -.025 |
| 4 | -.063 | .446 | .029 |
| 5 | -.045 | -.012 | .257 |
| 6 | .216 | .030 | -.027 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .038 | .187 | .137 |
| II | .339 | -.982 | .176 |
| III | .940 | -.020 | -.975 |

Matrix T. Direction Cosines of Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .969 | .946 | .980 |
| II | .181 | -.316 | .172 |
| III | .168 | .076 | -.098 |

Matrix TT' = Rpq. Cosines of Angles Between Primary Vectors $T_p$[b]

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | .872 | 1.001 | |
| C | .964 | .865 | 1.000 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

[b] Factor A should read Factor C
Factor C should read Factor A

## TABLE XII

## MATRICES OF OBLIQUE ROTATION OF AXES FOR FULL-BLOOD INDIAN CHILDREN, FEMALES, INTERMEDIATE GRADE LEVEL GROUP

Extended Factor Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.33333 | 1.000 | .379 | .104 |
| 2 | 1.25786 | 1.000 | .052 | .123 |
| 3 | 1.22549 | 1.000 | -.338 | -.048 |
| 4 | 1.32275 | 1.000 | -.440 | .026 |
| 5 | 1.22850 | 1.000 | -.039 | -.193 |
| 6 | 1.32979 | 1.000 | .356 | -.157 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .12 | .36 | .09 |
| II | .26 | -1.00 | .21 |
| III | 1.00 | - .03 | -1.00 |

Matrix $C = \Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda pq$

| Factor | A | B | C |
|---|---|---|---|
| A | .999 | | |
| B | -.223 | 1.001 | |
| C | -.876 | - .136 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | .232 | -.015 | .048 |
| 2 | .196 | .228 | -.017 |
| 3 | .075 | .537 | .053 |
| 4 | .023 | .569 | -.021 |
| 5 | -.065 | .310 | .218 |
| 6 | .040 | .006 | .236 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda p$

| Factor | A | B | C |
|---|---|---|---|
| I | .115 | .339 | .088 |
| II | .250 | -.941 | .205 |
| III | .961 | -.028 | -.975 |

Matrix T. Direction Cosines of Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .931 | .913 | .921 |
| II | .331 | -.408 | .337 |
| III | .153 | -.004 | -.197 |

Matrix TT' = Rpq. Cosines of Angles Between Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | .714 | 1.000 | |
| C | .939 | .704 | 1.001 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE XIII

## MATRICES OF OBLIQUE ROTATION OF AXES FOR FULL-BLOOD INDIAN CHILDREN, FEMALES, ADVANCED GRADE LEVEL GROUP

Extended Factor
Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.48148 | 1.000 | .382 | -.247 |
| 2 | 1.27877 | 1.000 | .225 | -.183 |
| 3 | 1.21951 | 1.000 | -.398 | -.102 |
| 4 | 1.37174 | 1.000 | -.494 | .196 |
| 5 | 1.30039 | 1.000 | .142 | .140 |
| 6 | 1.39860 | 1.000 | .341 | .267 |

Matrix S = L. Direction
Coefficients of a Normal
to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .10 | .30 | .19 |
| II | .35 | -1.00 | .22 |
| III | 1.00 | - .14 | -1.00 |

Matrix $C = \Lambda'\Lambda$ . Cosines of
Angles Between Reference
Vectors $\Lambda pq$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | -.410 | 1.000 | |
| C | -.817 | -.031 | .999 |

Rotated Oblique Factor
Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | -.009 | -.030 | .338 |
| 2 | -.003 | .075 | .317 |
| 3 | -.109 | .554 | .161 |
| 4 | .085 | .530 | -.081 |
| 5 | .210 | .101 | .059 |
| 6 | .327 | -.053 | -.002 |

Matrix $\Lambda$ . Direction Cosines
of Reference Vectors $\Lambda p$

| Factor | A | B | C |
|---|---|---|---|
| I | .094 | .285 | .182 |
| II | .329 | -.949 | .211 |
| III | .940 | -.133 | -.960 |

Matrix T. Direction Cosines of Primary
Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .940 | .890 | .930 |
| II | .249 | -.452 | .307 |
| III | .233 | .069 | -.201 |

Matrix $TT' = Rpq$. Cosines
of Angles Between Primary
Vectors $T_p$[b]

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | .740 | 1.001 | |
| C | .904 | .675 | 1.000 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

[b] Factor A should read Factor C
Factor C should read Factor A

## TABLE XIV

## MATRICES OF OBLIQUE ROTATION OF AXES FOR WHITE CHILDREN, MALES, ELEMENTARY GRADE LEVEL GROUP

Extended Factor Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.16414 | 1.000 | .204 | .198 |
| 2 | 1.17647 | 1.000 | .195 | .116 |
| 3 | 1.12740 | 1.000 | -.240 | .159 |
| 4 | 1.35318 | 1.000 | -.438 | -.115 |
| 5 | 1.27065 | 1.000 | .083 | -.126 |
| 6 | 1.33690 | 1.000 | .186 | -.151 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .10 | .17 | .13 |
| II | .36 | -1.00 | .19 |
| III | 1.00 | .17 | -1.00 |

Matrix $C = \Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda pq$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | - .157 | .999 | |
| C | - .839 | -.320 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | .299 | -.001 | -.024 |
| 2 | .229 | -.005 | .042 |
| 3 | .144 | .377 | -.064 |
| 4 | -.119 | .423 | -.049 |
| 5 | .003 | .050 | .208 |
| 6 | .011 | -.030 | .231 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda p$

| Factor | A | B | C |
|---|---|---|---|
| I | .094 | .165 | .127 |
| II | .337 | -.972 | .185 |
| III | .937 | .165 | -.975 |

Matrix T. Direction Cosines of Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .968 | .921 | .979 |
| II | .192 | -.387 | .141 |
| III | .162 | .046 | -.149 |

Matrix TT' = Rpq. Cosines of Angles Between Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | .825 | 1.000 | |
| C | .951 | .840 | 1.001 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE XV

## MATRICES OF OBLIQUE ROTATION OF AXES FOR WHITE CHILDREN, MALES, INTERMEDIATE GRADE LEVEL GROUP

Extended Factor Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.19474 | 1.000 | .399 | .111 |
| 2 | 1.15741 | 1.000 | .126 | .199 |
| 3 | 1.16822 | 1.000 | -.264 | .186 |
| 4 | 1.27551 | 1.000 | -.459 | -.033 |
| 5 | 1.24378 | 1.000 | -.031 | -.132 |
| 6 | 1.35318 | 1.000 | .192 | -.269 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .17 | .18 | .13 |
| II | .62 | -1.00 | .09 |
| III | 1.00 | .40 | -1.00 |

Matrix $C = \Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda_{pq}$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | - .147 | 1.000 | |
| C | - .766 | -.422 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | .372 | -.134 | .045 |
| 2 | .325 | .106 | -.050 |
| 3 | .138 | .406 | -.068 |
| 4 | -.098 | .450 | .094 |
| 5 | .013 | .117 | .205 |
| 6 | .012 | -.081 | .304 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .143 | .165 | .128 |
| II | .522 | -.916 | .089 |
| III | .841 | .366 | -.988 |

Matrix T. Direction Cosines of Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .962 | .919 | .972 |
| II | .232 | -.387 | .087 |
| III | .164 | .084 | -.219 |

Matrix $TT' = R_{pq}$. Cosines of Angles Between Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.001 | | |
| B | .807 | 1.001 | |
| C | .923 | .841 | 1.000 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE XVI

## MATRICES OF OBLIQUE ROTATION OF AXES FOR WHITE CHILDREN, MALES, ADVANCED GRADE LEVEL GROUP

Extended Factor Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.24070 | 1.000 | .423 | .257 |
| 2 | 1.27065 | 1.000 | .258 | .250 |
| 3 | 1.16959 | 1.000 | -.331 | .150 |
| 4 | 1.33333 | 1.000 | -.445 | .119 |
| 5 | 1.28205 | 1.000 | .065 | -.252 |
| 6 | 1.43062 | 1.000 | .239 | -.282 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .15 | .25 | .20 |
| II | .74 | -1.00 | .18 |
| III | 1.00 | .32 | -1.00 |

Matrix C = $\Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda pq$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | -.284 | 1.000 | |
| C | -.645 | -.402 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | .463 | -.068 | .015 |
| 2 | .372 | .053 | -.003 |
| 3 | .037 | .499 | -.008 |
| 4 | -.036 | .510 | .001 |
| 5 | .033 | .075 | .350 |
| 6 | .025 | -.051 | .354 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda p$

| Factor | A | B | C |
|---|---|---|---|
| I | .120 | .232 | .193 |
| II | .591 | -.927 | .174 |
| III | .798 | .296 | -.966 |

Matrix T. Direction Cosines of Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .921 | .928 | .953 |
| II | .307 | -.353 | .156 |
| III | .239 | .121 | -.258 |

Matrix TT' = Rpq. Cosines of Angles Between Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | .775 | 1.000 | |
| C | .864 | .798 | .999 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE XVII

## MATRICES OF OBLIQUE ROTATION OF AXES FOR WHITE CHILDREN, FEMALES, ELEMENTARY GRADE LEVEL GROUP

Extended Factor
Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.15741 | 1.000 | .209 | -.117 |
| 2 | 1.15473 | 1.000 | .090 | -.079 |
| 3 | 1.12233 | 1.000 | -.232 | .057 |
| 4 | 1.30890 | 1.000 | -.385 | .177 |
| 5 | 1.30548 | 1.000 | .304 | .287 |
| 6 | 1.32275 | 1.000 | .217 | -.022 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .01 | .21 | .20 |
| II | .40 | -1.00 | .20 |
| III | 1.00 | .32 | -1.00 |

Matrix $C = \Lambda'\Lambda$ . Cosines of Angles Between Reference Vectors $\Lambda_{pq}$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | -.067 | 1.000 | |
| C | -.821 | -.429 | .999 |

Rotated Oblique Factor
Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | .189 | .030 | .298 |
| 2 | -.026 | -.076 | .247 |
| 3 | -.021 | .385 | .082 |
| 4 | .023 | .467 | -.040 |
| 5 | .298 | -.004 | -.020 |
| 6 | .052 | -.012 | .193 |

Matrix $\Lambda$ . Direction Cosines of Reference Vectors $\Lambda_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .009 | .196 | .192 |
| II | .371 | -.934 | .192 |
| III | .929 | .299 | -.962 |

Matrix T. Direction Cosines of Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .932 | .937 | .981 |
| II | .273 | -.327 | .179 |
| III | .240 | .122 | -.081 |

Matrix TT' = Rpq. Cosines of Angles Between Primary Vectors $T_p$[b]

| Factor | A | B | C |
|---|---|---|---|
| A | 1.001 | | |
| B | .813 | 1.000 | |
| C | .944 | .851 | 1.001 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

[b] Factor A should read Factor C
Factor C should read Factor A

## TABLE XVIII

## MATRICES OF OBLIQUE ROTATION OF AXES FOR WHITE CHILDREN, FEMALES, INTERMEDIATE GRADE LEVEL GROUP

Extended Factor Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.19904 | 1.000 | .329 | .146 |
| 2 | 1.18483 | 1.000 | .108 | .226 |
| 3 | 1.17096 | 1.000 | -.288 | .128 |
| 4 | 1.22699 | 1.000 | -.369 | -.066 |
| 5 | 1.23305 | 1.000 | .011 | -.084 |
| 6 | 1.43267 | 1.000 | .248 | -.274 |

Matrix S = L. Direction Coefficients of a Normal to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .13 | .19 | .15 |
| II | .49 | -1.00 | .31 |
| III | 1.00 | .15 | -1.00 |

Matrix $C = \Lambda'\Lambda$. Cosines of Angles Between Reference Vectors $\Lambda pq$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | - .183 | 1.000 | |
| C | - .699 | - .397 | 1.000 |

Rotated Oblique Factor Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | .325 | -.094 | .083 |
| 2 | .308 | .086 | -.034 |
| 3 | .089 | .413 | -.054 |
| 4 | -.085 | .435 | .079 |
| 5 | .037 | .131 | .182 |
| 6 | -.014 | -.067 | .330 |

Matrix $\Lambda$. Direction Cosines of Reference Vectors $\Lambda p$

| Factor | A | B | C |
|---|---|---|---|
| I | .116 | .185 | .142 |
| II | .437 | -.972 | .293 |
| III | .892 | .146 | -.946 |

Matrix T. Direction Cosines of Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .954 | .943 | .968 |
| II | .213 | -.330 | .154 |
| III | .209 | .039 | -.202 |

Matrix TT' = Rpq. Cosines of Angles Between Primary Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| A | .999 | | |
| B | .837 | 1.000 | |
| C | .914 | .854 | 1.001 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

## TABLE XIX

## MATRICES OF OBLIQUE ROTATION OF AXES FOR WHITE CHILDREN, FEMALES, ADVANCED GRADE LEVEL GROUP

Extended Factor
Matrix $E_o$

| Test[a] | $D_e$ | I | II | III |
|---|---|---|---|---|
| 1 | 1.25628 | 1.000 | .366 | -.284 |
| 2 | 1.25313 | 1.000 | .165 | -.162 |
| 3 | 1.21359 | 1.000 | -.433 | -.118 |
| 4 | 1.32626 | 1.000 | -.517 | .171 |
| 5 | 1.25628 | 1.000 | .090 | .244 |
| 6 | 1.37363 | 1.000 | .372 | .207 |

Matrix S = L. Direction
Coefficients of a Normal
to Each Coordinate Plane

| Factor | A' | B' | C' |
|---|---|---|---|
| I | .12 | .26 | .18 |
| II | .34 | -1.00 | .26 |
| III | 1.00 | - .01 | -1.00 |

Matrix C = $\Lambda'\Lambda$. Cosines of
Angles Between Reference
Vectors $\Lambda_{pq}$

| Factor | A | B | C |
|---|---|---|---|
| A | 1.000 | | |
| B | - .291 | 1.000 | |
| C | - .799 | - .187 | 1.001 |

Rotated Oblique Factor
Matrix V

| Test[a] | A | B | C |
|---|---|---|---|
| 1 | -.030 | -.079 | .425 |
| 2 | .011 | .075 | .293 |
| 3 | -.112 | .554 | .146 |
| 4 | .082 | .566 | -.090 |
| 5 | .296 | .129 | -.030 |
| 6 | .311 | -.080 | .048 |

Matrix $\Lambda$. Direction Cosines
of Reference Vectors $\Lambda_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .113 | .252 | .172 |
| II | .320 | -.968 | .248 |
| III | .941 | -.010 | -.954 |

Matrix T. Direction Co-
sines of Primary
Vectors $T_p$

| Factor | A | B | C |
|---|---|---|---|
| I | .942 | .893 | .948 |
| II | .243 | -.447 | .248 |
| III | .233 | .045 | -.199 |

Matrix TT' = Rpq. Cosines
of Angles Between Primary
Vectors $T_p$[b]

| Factor | A | B | C |
|---|---|---|---|
| A | 1.001 | | |
| B | .743 | .999 | |
| C | .907 | .727 | 1.000 |

[a] Test.--1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

[b] Factor A should read Factor C
Factor C should read Factor A

## TABLE XX

## FACTOR PATTERNS AND FACTOR STRUCTURES BY RACE, SEX, AND GRADE LEVEL GROUPS AFTER PERMUTATED ROTATION

| Factor | Males | | | | | | | | | Females | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Elementary | | | Intermediate | | | Advanced | | | Elementary | | | Intermediate | | | Advanced | | |
| Test[a] | A | B | C | A | B | C | A | B | C | A[b] | B | C[b] | A | B | C | A[b] | B | C[b] |
| | Full-Blood Indian Children | | | | | | | | | | | | | | | | | |
| 1 | .42 | | | .52 | | | .65 | | | | | | .23 | | | .34 | | |
| 2 | .46 | | | .29 | | | .43 | | | .21 | | | .20 | .23 | | .32 | | |
| 3 | | -.37 | | | -.52 | | | -.53 | | | .39 | | | .54 | | | .49 | |
| 4 | | -.58 | | | -.56 | | | -.65 | | | .45 | | | .57 | | | .47 | |
| 5 | .39 | | | | -.27 | | | -.33 | | .26 | | | | .31 | .22 | | | .21 |
| 6 | .54 | | | .51 | | | | | | | | .22 | | | .24 | | | .33 |
| | White Children | | | | | | | | | | | | | | | | | |
| 1 | .30 | | | .37 | | | .46 | | | .30 | | | .33 | | | .43 | | |
| 2 | .23 | | | .33 | | | .37 | | | .25 | | | .31 | | | .29 | | |
| 3 | | .38 | | | .41 | | | .50 | | | .39 | | | .41 | | | .55 | |
| 4 | | .42 | | | .45 | | | .51 | | | .47 | | | .44 | | | .57 | |
| 5 | | | .21 | | | .21 | | | .35 | | | .30 | | | | | | .30 |
| 6 | | | .23 | | | .30 | | | .35 | | | | | | .33 | | | .31 |

[a] Test.—1. Reading Vocabulary, 2. Reading Comprehension, 3. Arithmetic Reasoning, 4. Arithmetic Fundamentals, 5. Language, 6. Spelling

[b] Factor A should read Factor C
Factor C should read Factor A

and X give the unrotated factor loadings, the rotated oblique factor loadings, the direction coefficients of the normal to each co-ordinate plane, the transformation matrix, and the matrix of cosines of the angular separations of the oblique axes for full-blooded Indian children at the elementary, intermediate, and advanced levels, respectively.

The direction coefficients of a normal to each co-ordinate plane, Matrix S, when normalized, resulted in the transformation Matrix Λ. The column numbers represent the direction cosines of the reference vectors with respect to the unrotated factor Matrix $F_o$. Post-multiplying the unrotated factor matrix by the transformation matrix yielded the rotated oblique factor Matrix V. The cosines of angles between the reference vectors, Matrix C (intercorrelations) were obtained by multiplying the matrix of direction cosines of reference vectors by its transformation.

After the oblique rotations were accomplished for all the other groups, the cosines of angles between the reference vectors all had negative signs indicating a positive correlation between factors A and B, factors A and C, and factors B and C. The cosines between the reference vectors are shown in Matrix C of Tables VIII through XIX, and the cosines between the primary factors are shown in Matrix $R_{pq}$ of Tables XI through XIX. The intercorrelations were computed by using

$$T = D\Lambda^{-1} \quad \text{and} \quad R_{pq} = D\Lambda^{-1}(D\Lambda^{-1})' = TT'$$

where T is the inverse of Matrix Λ normalized or the factor matrix for the primary vectors and $R_{pq}$ is the correlation between the primary factors which is obtained by matrix multiplication of Matrix T by its transpose. The rotated oblique factor Matrix V for these groups was determined by post-multiplying the unrotated factor Matrix $F_o$ of Tables V, VI, and VII by the respective transformation Matrix Λ of Tables XI through XIX.

Matrix C in Tables VIII through XIX show the intercorrelations between the reference vectors while Matrix $R_{pq}$ in Tables XI through XIX show intercorrelations between the primary vectors. These correlations were all significantly different from zero. This indicates that the factors were not independent and that a relationship existed among them. The highest correlation existed between factors A and C which had significant loadings on Tests 1, 2, 5, and 6. In all cases, Tests 3 and 4 had significant loadings of Factor B which was also significantly related to the other two factors. These relationships indicated the presence of a second-order factor common to all the tests.

In order to facilitate comparison of the various groups, factor patterns of the rotated oblique factor loadings on each of the six tests are presented in Table XX. Factor loadings of less than .20 were not included in the pattern plan. The elementary and advanced grade level groups for females of both races showed patterns in a reverse direction. Permutated rotation was used to reverse the pattern so that comparisons of the sets of factor patterns could be made more readily (55). Factors A and C in Matrix V in Tables XI, XIII, XVII, and XIX, for these two grade level groups, were interchanged by post-multiplying by the permutation matrix. Correspondingly, the interchange in Matrix $R_{pq}$ was obtained by pre-multiplying and post-multiplying it by the permutation matrix which resulted in an interchange of $r_{AB}$ and $r_{CB}$. This reversal did not effect the values of the factor loadings.

Two significant factors were present for full-blooded Indian male children at each grade level, while three significant factors were present for the other groups.

Rather consistent patterns of factors were evident among grade level groups for full-blooded Indian male children on all tests except 5 and 6. Test 5 changed from a Factor A to a Factor B loading at the intermediate and advanced levels, while Factor A vanished on Test 6 at the advanced level. For full-blooded Indian female children a rather similar pattern existed among the three grade level groups. The "staircase effect" was present at all levels with the elementary group showing a Factor C loading on Test 5 and no significant loading on Test 1. A similar pattern was noted at the intermediate level with Factor B loadings on Tests 2 and 5. For white male and female children at all grade levels, three groups of two tests each, that is, Tests 1 and 2, 3 and 4, and 5 and 6 had significant loadings except for females on Test 5 at the intermediate level and on Test 6 at the elementary level.

The largest factor loadings were evident for full-blooded Indian male children. In most cases the loadings tended to increase by grade level groups for each race and sex. The sets of factor patterns between race groups appeared to be more varied than between sex. This was noted especially for females where the factors clustered in groups of two on the same tests at the various levels for white children. The same pattern was noted at the advanced level for full-blooded Indian children, but at the other two levels greater pattern variability was present. In relation to sex, the factor patterns were similar for white children, but they showed more inconsistency for the full-blooded Indian children.

## III. Interpretation

Since the tests used in this study were all paper-and-pencil variety, and since the tests purported to measure the basic educational skills, the nature of the factors were considered in view of educational achievement. This limitation was undesirable, but restrictive because no other measurements were available for the subjects studied.

In view of the tests' content, Factors A and C were identified as verbal factors and Factor B was identified as a numerical factor with a common factor underlying all the tests, which might be identified as a Schooling factor. It would be inappropriate to attempt an interpretation of the abilities necessary to perform the tasks called for by the various tests. The tests with more verbal material, Reading vocabulary, Reading comprehension, Language, and Spelling did not remain clustered as well as did those tests with number material, Arithmetic reasoning, and Arithmetic fundamentals.

### Evaluation of Hypotheses[2]

Hypothesis 1—A comparison between the mean raw scores computed by the critical ratio showed a significant difference, .05 level of confidence, on all tests for the three grade level groups in favor of white children. With regard to sex differences, the mean raw scores for female white children were significantly higher, .05 level of confidence, than for male white children with one exception: At the advanced grade level males surpassed females on the Arithmetic reasoning test. In the majority of cases, the mean raw scores were significantly higher, .05 level of confidence, for full-blooded Indian female children. The mean differences in favor of full-blooded Indian male children were on the Reading comprehension and Arithmetic reasoning tests at the elementary level, Reading vocabulary, Reading comprehension, and Arithmetic fundamentals at the intermediate level, and Reading vocabulary and Reading comprehension at the advanced level.

Hypothesis 2—Two factors were found for full-blooded Indian male children at the elementary, intermediate, and advanced grade levels. These were identified as a verbal and a numerical factor. The high correlations between these factors suggested a second-order general factor. Three factors were found for full-blooded Indian female children and for male and female white children at each of the three grade level groups.

Hypothesis 3—The factor loadings ranged from .20 to .65 for full-blooded Indian children and from .21 to .57 for white children. More high loadings appeared for full-blooded Indian male children than for any other race-sex group.

Hypothesis 4—The factor loadings on the reading tests were higher for full-blooded Indian male children than for white male children. In general, the differences in loadings were .10 or greater in favor of the full-blooded Indian children. A similar difference was evident on the language and spelling tests.

Hypothesis 5—Full-blooded Indian male children showed higher loadings of a number factor on the arithmetic tests than did white male children. Generally, the size of the loadings of this factor increased with increase in grade level. For full-blooded Indian female children the loadings were highest at the intermediate level and lowest at the elementary level, whereas, for the white female children the loadings were highest at the advanced level and approximately of the same magnitude at the elementary and intermediate level with moderate loadings.

Hypothesis 6—The correlation matrices for all groups analyzed showed highly significant positive relationships. This suggested an underlying common factor. The common factor was substantiated by the significant intercorrelations existing between the reference and the primary vectors.

## IV. Discussion

There is little evidence regarding differences in sets of factor patterns and factor structures by race and sex for the elementary, intermediate, and advanced grade level groups. Although the results did not conform ideally to Garrett's (25) differential hypothesis whereby abstract intelligence changes in organization from a general ability to an organized group of abilities or factors as age increases, the arithmetic tests showed rather consistent increases in factor loadings between the grade levels from the elementary to the advanced level. A similar trend was noted for both full-blooded Indian and white children on the reading tests which was more pronounced for males. This evidence was based on cross-sectional rather than longitudinal data. When the correlations between the reference vectors and the primary vectors are considered, in most cases the correlations decreased with increase in grade level. These correlations were significantly different from each other which indicated differentiation between grade level groups.

With regard to the interpretation of the factors it should be pointed out that with only six variables the reference frames would not be too stable, therefore, the factor loadings might be subject to change from sample to sample. It was difficult to adequately establish the traces on the test projections with only six tests and criterion tests with established factors for the subjects in this study were not available.

The high positive correlations among the tests and among the primary vectors strongly suggested the presence of a factor in the second-order domain. Apparently, this accounted for the low loadings on the first-order factors. It may be assumed that the tests are rather complex in that they not only involve a specific factor, but also the common factor defined by the tests. Tests 1, 2, 5, and 6 appeared to be more complex than Tests 3 and 4. The reading and language tests involved two factors in addition to the common factor. For full-blooded Indian male children the language test clustered with the arithmetic tests at the intermediate and advanced levels. Perhaps at these levels the three tests have some associative characteristics. A similar trend was noted for full-blooded Indian female children at the intermediate level with the reading comprehension test also joining the cluster. The clustering of the reading comprehension and language tests with the arithmetic tests did not occur for the other groups.

Since full-blooded Indian children and white children receive their education in American schools, it seems reasonable to assume that the Indian children should learn the same basic skills as the white children. The results as shown in this study as well as in other studies cited herein, indicated that significant differences existed between races in educational achievement. Apparently, factors other than school experiences play an important role. These factors could be genetic, environmental, and experiential. According to French (21), the factors might readily be conceptualized as experiential factors since they were obtained from measures on achievement tests. French (21) named the common factor which involves skills acquired in school, e.g., fractions, decimals, arithmetic problems, and vocabulary tests, as Schooling. This factor usually appears in the second-order domain. He calls attention to the interpretation of the general, Number, and Verbal factors which are all quite distinct, and the relation between the Verbal or Number factor has nothing to do with the actual mechanism of Verbal or Number ability. It is perhaps an arbitrary requirement of American schools which is determined by differences in amount of formal education. The findings in this study seem to corroborate French's interpretation.

Since the factor patterns and size of factor loadings are essentially the same for both racial groups by grade level and sex, it indicates that a large proportion of the content of the California Achievement Test Battery measures approximately the same basic educational skills at each grade level. This evidence of continuity of content and the variability of mean achievement suggests that various other factors are interferring with good articulation between elementary and intermediate and between intermediate and advanced grade levels for the two racial groups.

The significant differences in mean achievement and the slight discrepancies in the sets of factor patterns between full-blooded Indian and white children might be speculated from a theoretical viewpoint. The groups studied were reasonably comparable with respect to geographical location and educational grade placement. Assuming that these conditions were sufficiently controlled, it seems that other factors would account for the differences obtained. American culture perhaps determines very largely that both Indian children and white children develop basic educational skills since these skills are important to both races outside an educational setting. However, the two races undoubtedly place different values on these fundamental skills even though they both function in our American society. Because of a difference in the interaction between the individual's educational environment and his cultural milieu it seems reasonable to assume that Indian children are not as cognizant of the ultimate need for these basic educational skills. Other than cultural differences, slow work habits, the nature of the test items, poor motivation, level of aspiration, and lack of interest may also account for some of the differences encountered. The question also arises as to whether the tests were equally fair to both groups. Undoubtedly, these conditions could have effected the results of this study.

## V. Summary

The University of Kansas had conducted a fall achievement testing program for Indian and white children at the elementary, intermediate, and advanced grade levels for the United States Bureau of Indian Affairs during the period 1951-55. The California Achievement Test Batteries were used to obtain scores of educational achievement in the following basic skills: reading vocabulary, reading comprehension, arithmetic reasoning, arithmetic fundamentals, language, and spelling. When the mean scores of these tests were compared as to race, significant differences were found in favor of white children.

The sample was limited to all full-blooded Indian children and white children on whom a complete set of test scores was available. The distribution of the 14,888 subjects used in this study consisted of 5502 full-blooded Indian children and 9386 white children. The geographical locations of the administrative areas of the United States Bureau of Indian Affairs in which the testing was conducted were: 1) Phoenix, 2) Albuquerque, 3) Aberdeen, 4) Billings, 5) Muskogee, and 6) Anadarko.

Pearson product-moment correlations between the tests were computed for the various grade level groups by race and sex. The twelve intercorrelation matrices were factor analyzed by Thurstone's centroid method and the factors were subjected to oblique rotation by means of extended vectors in accordance with the concept of simple structure. Three factors resulted. With oblique rotation, the third factor was not significant for full-blooded Indian male children. Correlations between the reference vectors and the primary vectors were obtained. Factors A and C were identified as verbal factors and Factor B as a numerical factor with highly significant correlations between the factors. This indicated a second-order factor common to all tests.

The purpose of this study was to compare the sets of factor patterns and factor structures by race, sex, and grade levels. Conclusions from the factor analyses follow:

1. Two factors were found for full-blooded Indian male children at each of the three grade level groups, while three factors were found at each grade level for full-blooded Indian female children and for male and female white children.
2. Higher factor loadings were found for full-blooded Indian male children than for any other race-sex group. The factor loadings ranged from .20 to .65 for full-blooded Indian children and from .21 to .57 for white children.
3. In most cases the factor loadings of a verbal nature were higher for full-blooded Indian male children.
4. In general, the factor loadings of a numerical nature were higher for full-blooded Indian children except on the arithmetic reasoning test where white female children at the advanced level had the highest loading.
5. The sets of factor patterns for all groups were more similar than they were different.
6. A common factor of the second-order domain was evident throughout all groups. This factor is probably *Schooling*.

## FOOTNOTES

* This monograph is based on the author's dissertation completed at the University of Kansas. The data used in this study were obtained from the Indian Education Research Project conducted by the United States Bureau of Indian Affairs and the University of Kansas during the period 1951-55. Mr. L. M. Coombs, Education Specialist, United States Bureau of Indian Affairs, offered suggestions which were essential to the study. Dr. K. E. Anderson and Dr. A. H. Turney, Committee Members, read and evaluated the manuscript. Dr. R. N. Bradt, Department of Mathematics, provided helpful advice on the statistical and factorial analyses. Mr. C. Bradrick, Director of the Statistical Bureau, compiled the data for computation of the correlations. The author is indebted to these persons for their cooperation and assistance. The author is deeply indebted to his major advisor, Dr. E. Gordon Collister, not only for his excellent direction and advice, but also for his encouragement and inspiration.

** Now at the College of Education, University of Oklahoma.

1. The writer is indebted to the staff of the Statistical Bureau, University of Kansas, for carrying out the necessary machine operations for these factor analyses.

2. The following scale was used for the evaluation of factor loadings:

| | |
|---|---|
| Insignificant loadings | .00 to .20 |
| Low loadings | .20 to .30 |
| Moderate loadings | .30 to .50 |
| High loadings | .50 and above |

## BIBLIOGRAPHY

1. Anastasi, Anne. "Further Studies on the Memory Factor," *Archives of Psychology*, XXII (1932), pp. 1-60.
2. Anderson, K. E. *The Relative Achievements of the Objectives of Secondary School Science in a Representative Sampling of Fifty-Six Minnesota Schools*, unpublished Ph.D. dissertation, University of Minnesota, 1949.
3. Anderson, K. E., and others. *The Educational Achievement of Indian Children. A Re-examination of the Question: How Well are Indian Children Educated?* (Lawrence, Kansas: Haskell Institute, 1953).
4. Asch, S. E. "A Study of Change in Mental Organization," *Archives of Psychology*, XXVIII (1936), pp. 1-30.
5. Balinsky, B. "An Analysis of the Mental Factors of Various Age Groups from 9-60," *Genetic Psychology Monographs*, XXIII (1941) pp. 191-234.
6. Blackwell, A. M. "A Comprehensive Investigation into the Factors Involved in Mathematical Ability of Boys and Girls," *British Journal of Educational Psychology*, X (1940), pp. 143-53, 212-22.
7. Bryan, A. I. "Organization of Memory in Young Children," *Archives of Psychology*, XXIV (1934), pp. 1-56.
8. Carroll, J. B. "A Factor Analysis of Verbal Abilities," *Psychometrika*, VI (1941), pp. 279-307.
9. Clark, M. P. "Changes in Primary Mental Abilities with Age," *Archives of Psychology*, XLI (1944), pp. 1-30.
10. Clark, W. *A Study of Factors Related to Mastery of Skills in Reading, Arithmetic, and Written Expression in Los Angeles County Senior High Schools*, unpublished Ed.D. dissertation, University of Southern California, 1941.
11. Coleman, J. C. *California Test of Mental Maturity: Summary of Investigations Number Three* (Los Angeles: California Test Bureau, 1956), pp. 1-30.

12. Coombs, C. H. “A Factorial Study of Number Ability,” Psychometrika, VI (1941), pp. 161-89.
13. Coombs, L. M., and others. The Indian Child Goes to School (Lawrence, Kansas: Haskell Institute, 1956).
14. Crysostom, F. “L’organization des Traits Mentaux avec l’Age,” Annales de l’Assoc., Canadienne-pour l’Advancement des Sciences, IV (1938), pp. 1-124.
15. Cureton, E. E. “The Verbal Relations Factor and Vocabulary,” American Psychologist, II (1947), pp. 286-87.
16. Cureton, E. E. “Vocabulary as an Arbitrary Definer of the Verbal Relations Factor,” Invitational Conference on Testing Problems (New York: Educational Testing Service, 1948), pp. 47-9.
17. Curtis, H. A. “A Study of the Relative Effects of Age and Test Difficulty Upon Factor Patterns,” Genetic Psychology Monograph, XL (1949), pp. 99-148.
18. Davis, F. B. “Fundamental Factors of Comprehension in Reading,” Psychometrika, IX (1944) pp. 185-97.
19. Doppelt, J. E. “The Organization of Mental Abilities in the Age Range Thirteen to Seventeen,” American Psychologist, V (1949), p. 242.
20. Fitzgerald, J. A., and Ludeman, W. W. “The Intelligence of Indian Children,” Journal of Comparative Psychology, VI (1926), pp. 318-28.
21. French, J. W. “The Description of Aptitude and Achievement Tests in Terms of Rotated Factors,” Psychometric Monograph (Chicago: University of Chicago Press, 1951).
22. Fruchter, B. Introduction to Factor Analysis (New York: Van Nostrand Publishing Co., 1954).
23. Garrett, H. E., and others. “The Age Factor in Mental Organization,” Archives of Psychology, XXVI (1935), pp. 1-31.
24. Garrett, H. E. “Differentiable Mental Traits,” Psychological Record, II (1938), pp. 259-98.
25. Garrett, H. E. “A Developmental Theory of Intelligence,” American Psychologist, I (1946), p. 373.
26. Garth, T. R. “The Results of Some Tests on Full and Mixed Blood Indians,” Journal of Applied Psychology, V (1921), pp. 359-72.
27. Garth, T. R. “A Comparison of Mental Abilities of Mixed and Full Blood Indians on a Basis of Education,” Psychological Review XXIX (1922), pp. 221-36.
28. Garth, T. R. “A Comparison of the Intelligence of Mexican and Mixed and Full Blood Indian Children,” Psychological Review, XXX (1923), pp. 388-401.
29. Garth, T. R. “The Intelligence of Full Blood Indians,” Journal of Applied Psychology, IX (1925), pp. 382-89.
30. Garth, T. R. “The Intelligence of Mixed Blood Indians,” Psychological Bulletin, XXIV (1927), pp. 179-80.
31. Guilford, J. P. Psychometric Methods (New York: McGraw Hill, 1954).
32. Hall, W. E., and Robinson, F. F. “An Analytical Approach to the Study of Reading Skills,” Journal of Educational Psychology, XXXVI (1945), pp. 429-42.
33. Harris, C. W. “An Exploration of Language Skill Patterns,” Journal of Educational Psychology, XXXIX (1948), pp. 321-36.
34. Helmer, Velma. The American Indian and Mental Tests, unpublished M. A. thesis, University of Kansas, 1925.
35. Holzinger, K. J., and Swineford, Frances. “A Study in Factor Analysis: The Stability of a Bi-factor Solution,” Supplemented Educational Monograph, No. 48 (Chicago: University of Chicago Press, 1939), pp. 1-91.
36. Hunter, W. S. ans Sommermier, E. “The Relation of Degree of Indian Blood to Score on the Otis Intelligence Test,” Journal of Comparative Psychology, II (1922), pp. 257-77.
37. Joseph, Alice, and others. The Desert People: A Study of the Papago Indians (Chicago: University of Chicago Press, 1949).
38. Kelley, T. L. Crossroads in the Mind of Man: A Study of Differentiable Mental Abilities (Los Angeles: Stanford University Press, 1928).
39. Klineberg, O. An Experimental Study of Speed and Other Factors in Racial Differences,” Archives of Psychology, XV, No. 93 (1938), pp. 1-111.
40. Leighton, Dorothea C., and Kluckhohn, C. Children of the People (Cambridge: Harvard University Press, 1947).
41. MacGregor, G. Warriors Without Weapons: A Study of the Society and Personality Development of the Pine Ridge Sioux (Chicago: University of Chicago Press, 1945).
42. Meriam, L., and others. The Problem of Indian Administration (Baltimore: Johns Hopkins Press, 1928).
43. Peterson, S. How Well are Indian Children Educated? (Lawrence, Kansas: Haskell Institute, 1948).
44. Reichard, S. “Mental Organization and Age Level, Archives of Psychology, XLI (1944), pp. 1-29.
45. Richards, S. “Genetic Emergence of Factor Specificity,” Psychometrika, VI (1941), pp. 37-42.
46. Rowe, E. C. “Five Hundred Sixty-Seven White and Two Hundred Sixty-Eight Indian Children Tested by the Binet-Simon Tests,”

Pedogogical Seminary, XXI (1914), pp. 454-68.

47. Schiller, B. "Verbal, Numerical, and Spatial Abilities of Young Children," Archives of Psychology, XXIV (1934), pp. 1-69.

48. Schneck, M. M. R. "The Measurement of Verbal and Numerical Abilities," Archives of Psychology, XVII (1929), pp. 1-69.

49. Swineford, Frances. "Growth in General and Verbal Bi-factors from Grades VII-X," Journal of Educational Psychology, XXXVIII (1947), pp. 257-72.

50. Swineford, Frances. "The Nature of the General, Verbal, and Spatial Bi-factors," Supplemental Education Monograph (Chicago: University of Chicago Press, 1948), pp. 1-72.

51. Swineford, Frances. "General Verbal and Spatial Bi-factors After Three Years," Journal of Educational Psychology, XL (1949), pp. 353-60.

52. Thompson, Laura and Joseph, Alice. The Hopi Way (Chicago: University of Chicago Press, 1944).

53. Thurstone, L. L. "Note on a Reanalysis of Davis' Reading Tests," Psychometrika, XI (1946), pp. 185-88.

54. Thurstone, L. L. Multiple-Factor Analysis (Chicago: University of Chicago Press 1947).

55. Wrigley, C., and Neuhaus, J. O. The Matching of Two Sets of Factors, mimeographed paper presented to the American Psychological Association, San Francisco, September 5, 1955.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# ATTITUDE OF KINDERGARTEN CHILDREN TOWARD SPECIALLY PREPARED STORY MATERIALS*

BILL SNIDER**
State University of Iowa

THE DEVELOPMENT of materials for helping the child become sensitive to the dynamics of behavior has been one of the aims of the Preventive Psychiatry Research Program in the Child Welfare Research Station at the State University of Iowa. Among the material prepared is a group of narratives designed to help the kindergarten child develop a causal orientation toward his social environment. If these stories are to be an effective medium for introducing the young child to a causal way of thinking, it seems necessary that the children should enjoy listening to them. The primary purpose of this study, therefore, was to discover how well kindergarten children liked to listen to and discuss causally oriented narratives specially written for their age. The statistical procedure chosen to determine this was the method of paired comparisons.

The purpose of the causally oriented story is to introduce the child to the differences between a non-causal or surface approach and a causal approach to human situations. In the narrative, a social situation is presented and someone begins to act in a non-causal way. The problem is then re-examined and the individual proceeds more causally after which the effects of this type of action are considered The reading of the story is accompanied by a discussion of differences between the two approaches and some of the probable effects of each.

## Description of the Investigation

During the school year 1954-55, a number of teachers participated in the Preventive Psychiatry Program (5) by presenting causally oriented materials to their classes. The teachers had taken part in an intensified training program through summer workshops set up to train them to practice and teach the dynamic approach. The pupils in two of the kindergarten groups were chosen as subjects for this study. Class A consisted of 41 children in the morning and afternoon classes of teacher A, and Class B of 61 pupils in the morning and afternoon classes of teacher B.

In the method of paired comparisons, the subject is presented a number of different stimuli in pairs and is asked to judge which is greater with respect to some characteristic, such as, which of two objects is heavier, which is prettier, or which of two activities he prefers. In this particular investigation each child was asked which of two kindergarten activities he preferred. By statistical analysis of paired comparisons it is possible to assign a set of numbers to the different stimuli the subjects have been asked to judge.

Six different kindergarten activities were studied, namely: having music, working puzzles, having the teacher read ordinary or regular kindergarten stories, playing house, painting at the easel, and having the teacher read the causally oriented primary stories. These kindergarten activities were combined in all possible pairs. Each child was taken privately by a trained person[1] and asked which of each pair of activities he preferred.

The first few minutes with each child was spent in establishing rapport and seeing if the child understood the difference between the two types of stories being used as stimuli. The regular kindergarten stories, often read by the teacher, consisted of a pile of several books from the book table. The causally oriented stories were designated as the book without any pictures from which the teacher read to them; a copy of this book was on the table so that the child could recall it. A final check was made by asking the child if he knew the difference between the stories in this book without any pictures in it and the stories here—pointing to the books from the table. The administrator did not let the child see the books from the table other than the stack as the child was to respond to the books from the library table rather than to specific stories, and, likewise, the book containing the causally oriented stories was only shown to the child without permitting him to leaf through it.

---

* The preparation of this paper was supported by the Grant Foundation.

** Address: Preventive Psychiatry Research Program, Child Welfare Research Station.

[1] Mrs. Helen Marie Clark, Elementary Supervisor for the Preventive Psychiatry Program, 1955-56, administered the test.

The presentation of the pairs of stimuli were presented in such a manner as to adhere to the following principles as closely as possible (10, 11):

1. Maintain the greatest possible spacing between pairs involving any given member of the stimulus group.
2. Present a stimulus equally often first and second. This is referred to as taking account of "space errors."
3. Present stimuli as nearly in the same manner as is possible so as to avoid suggestion effects.

In order to insure the fullfillment of the third principle, the question for the presentation of each pair of objects was carefully planned and written in the instructions. Examples of these questions were:

1. Which would you rather do, listen to a story from this book (the causally oriented stories) or listen to a story from this book (regular stories)? (In each case only pointing to the book or books.)
2. Which would you rather do, have music time when you sing or work with a puzzle?
3. Which would you rather do, play house or paint at the easel?

## Analyses of the Data

The first step in the analysis of preferences of the six kindergarten activities used in this study was the preparation of a proportion matrix showing the proportion of times each activity was preferred over its paired activity. Table I and Table II give the proportion matrices for the children in Classes A and B respectively. In these matrices each stimulus at the top is judged greater than each stimulus on the side by the corresponding proportion. For example, in Table I the proportion of the time "Music" was preferred over "Easel Painting" is .463. In Table II the proportion of time "Playing House" was preferred over "Regular Stories" is .623.

The analysis of the proportion matrices in Tables I and II is based upon the application of Thurstone's law of comparative judgment (13). This law is stated in the following form:

$$R_j - R_k = z_{jk}\sqrt{s_j^2 + s_k^2 - 2r_{jk}s_js_k}$$

where $R_j$ and $R_k$ = mean psychological values characteristically attached to stimuli $S_j$ and $S_k$ respectively,
$z_{jk}$ = standard measure distance or deviate from the mean of a unit normal distribution,
$s_j$ and $s_k$ = standard deviations of distributions $R_{hj}$ and $R_{hk}$ respectively, and
$r_{jk}$ = coefficient of correlation between $R_{hj}$ and $R_{hk}$.

The radical term is the standard deviation of the differences $R_{hj} - R_{hk}$ is expressed as a function of the terms under the radical.

In Thurstone's Case III, he assumes $r_{jk} = 0$, that there is no correlation between responses to any pair of stimuli. Under this assumption the law of comparative judgment reduces to the form

$$R_j - R_k = z_{jk}\sqrt{s_j^2 + s_k^2}$$

Scale-values in terms of preference for the six activities as judged by the children in Class A and Class B were analyzed under the assumption of Thurstone's Case III. This was done for Class A and Class B separately and for the combined group.

We find the $z_{jk}$ values from a knowledge of the experimental proportions $p_{j>k}$. The $z_{jk}$ is the z-value, given in a table of the normal curve, for the respective proportion, consider the proportion $p_{j>k}$ as the area under the normal curve. When the proportion p is greater than .500 it's z-value is positive; when p is less than .500 it's z-value is negative.

Table III gives the $z_{jk}$ values for the respective proportions obtained for Class A as given in Table I. The sum column in Table III is used as a means of checking the computations. The sum of the sums should equal zero, and the sum of the column means should also equal zero.

Burros (1) proposed some convenient formulas for estimating the relative values for the standard deviations:

$$s_j = \frac{c}{\Sigma V_{z_j}}$$

where $s_j$ = the standard deviation for the dispersion of stimulus $S_j$
$V_{z_j}$ = the variance of deviations $z_{jk}$ in each column of Matrix $Z_5$ from the column mean, $M_{z_{jk}}$, these values are found in Table III for Class A. )

c is estimated by the formula

$$c \doteq \frac{n}{\Sigma \frac{1}{V_{z_j}}}$$

TABLE I

PROPORTION MATRIX FOR THE SIX KINDERGARTEN ACTIVITIES IN TERMS OF PREFERENCE AS JUDGED BY THE FORTY-ONE PUPILS IN CLASS A* (MATRIX $P_A$)

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories |
|---|---|---|---|---|---|---|
| Regular Stories | .500 | .561 | .561 | .537 | .610 | .902 |
| Music | .439 | .500 | .707 | .561 | .537 | .732 |
| Playing House | .439 | .293 | .500 | .659 | .561 | .780 |
| Working Puzzles | .463 | .439 | .341 | .500 | .659 | .683 |
| Easel Painting | .390 | .463 | .439 | .341 | .500 | .659 |
| Causal Stories | .098 | .268 | .220 | .317 | .341 | .500 |
| $\Sigma p_j > k$ | 2.329 | 2.524 | 2.768 | 2.915 | 3.208 | 4.256 |

*Kindergarten Activities have been arranged in order of increasing $\Sigma p_j > k$

TABLE II

PROPORTION MATRIX FOR THE SIX KINDERGARTEN ACTIVITIES IN TERMS OF PREFERENCE AS JUDGED BY THE SIXTY-ONE PUPILS IN CLASS B* (MATRIX $P_B$)

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories |
|---|---|---|---|---|---|---|
| Regular Stories | .500 | .574 | .623 | .656 | .787 | .918 |
| Music | .426 | .500 | .459 | .459 | .656 | .705 |
| Playing House | .377 | .541 | .500 | .607 | .525 | .623 |
| Working Puzzles | .344 | .541 | .393 | .500 | .627 | .623 |
| Easel Painting | .213 | .344 | .475 | .328 | .500 | .443 |
| Causal Stories | .082 | .295 | .377 | .377 | .557 | .500 |
| $\Sigma p_j > k$ | 1.942 | 2.795 | 2.827 | 2.927 | 3.697 | 3.812 |

*Kindergarten Activities have been arranged in order of increasing $\Sigma p_j > k$ across the columns.

TABLE III

Z-VALUES FOR THE RESPECTIVE PROPORTIONS OF THE SIX KINDERGARTEN ACTIVITIES IN TERMS OF PREFERENCE AS JUDGED BY THE FORTY-ONE PUPILS IN CLASS A (MATRIX $Z_{5A}$)

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories | Sum |
|---|---|---|---|---|---|---|---|
| Regular Stories | .000 | .154 | .154 | .093 | .279 | 1.293 | 1.973 |
| Music | - .154 | .000 | .545 | .154 | .093 | .619 | 1.257 |
| Playing House | - .154 | - .545 | .000 | .410 | .154 | .772 | .637 |
| Working Puzzles | - .093 | - .154 | - .410 | .000 | .410 | ..476 | .229 |
| Easel Painting | - .279 | - .093 | - .154 | - .410 | .000 | .410 | - .526 |
| Causal Stories | -1.293 | - .619 | - .772 | - .476 | - .410 | .000 | -3.570 |
| Sum $z_{jk}$ | -1.973 | -1.257 | - .637 | - .229 | .526 | 3.570 | .000 |
| Mean $z_{jk}$ | - .329 | - .210 | - .106 | - .038 | .088 | .595 | .000 |

TABLE IV

ESTIMATION OF THE STANDARD DEVIATIONS OF THE RESPONSES TO THE SIX KINDERGARTEN ACTIVITIES AS JUDGED BY THE FORTY-ONE KINDERGARTEN CHILDREN IN CLASS A

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories | Sum |
|---|---|---|---|---|---|---|---|
| $V_{z_j}$ | .1928 | .0788 | .1735 | .0977 | .0667 | .1536 | |
| $1/V_{z_j}$ | 5.187 | 12.690 | 5.764 | 10.235 | 14.993 | 6.510 | 55.379 |
| $s_j$ | .562 | 1.374 | .624 | 1.108 | 1.624 | .705 | |
| $s^2_j$ | .316 | 1.888 | .389 | 1.228 | 2.637 | .497 | 5.996 |

where n is the number of stimuli, and $V_{z_j}$ is defined as above.

Substituting the $z_{jk}$ and $M_{z_{jk}}$ values in Table III in the above formulas we can estimate the standard deviations for the six kindergarten activities as judged by the forty-one children in Class A. Table IV gives the statistics needed for computing Burro's formula for the Class A group.

The sum of the reciprocals of the variances is 55.379 and the number of stimuli is six, giving a c-value of .1083. As a check on the computations the $s_j$ values should sum to n, the number of stimuli. This is due to the fact that this method assures a mean $s_j$ of 1.000.

Table V contains the $(R_j - R_k)$ values which are computed according to the assumptions of Thurstone's Case III, $R_j - R_k = z_{jk}\sqrt{s_j^2 + s_k^2}$ and the $R_j$ scale values are based upon these assumptions. The $R_j$ value is a method of arbitrarily assigning the zero point to the lowest mean and adding the amount necessary to make the lowest mean .000 to each of the mean values. Either the mean values or the $R_j$ values can be taken as scale values for the different kindergarten activities. Guilford (2) presents a method of locating a psychological meaningful zero point; however, this requires addtional data.

The sum column in Table V is used as a means of checking the computations in the table, the sum of the sums should equal zero, the sum of the columns means should also equal or approximate zero, the sum of the $R_j$'s should equal n times the number added to each mean $M_{z_{jk}}$, or .353n, where n is the number of stimuli.

Given the scale values in Table V, the next problem is to determine how well the observed proportion of pupils' judgment agree with the expected proportion as determined by the above scale values. This test of internal consistency is made by utilizing Mosteller's method (8) by which a single chi-sqare test is made for the obtained and expected proportions. The first step is to find the expected $z'_{jk}$ values. These are obtained by applying the following formula:

$$z'_{jk} = \frac{R_j - R_k}{\sqrt{s_j^2 + s_k^2}}$$

where the $R_j - R_k$ values in the numerator are obtained from the interpair scale differences given in the $R_j$ row in Table V. The denominator is the square root of the sum of the paired $s_j^2$'s in the last row of Table IV. These $z'_{jk}$ values are given in Table VI. Only half the matrix is needed since the two halves are identical in absolute values. The diagonal cells have been ignored as there are no actual experimentally obtained values.

The expected proportions p', given in Table VII, are the areas under the normal curve for the respective $z'_{jk}$ values in Table VI. If z' should be negative the area under the smaller portion of the curve is used, whereas when z' is positive the area under the larger portion is used.

The test for internal consistency as proposed by Mosteller (8) does not use the actual p and p' values since many of the proportions might be close to 1.00, which is the region where sampling distributions of proportions are not normal even for reasonably large samples. In order to compensate for this lack of normality, each proportion is transformed to a statistic θ, whose sampling distribution is normal. Theta being the angle whose sine is the $\sqrt{p}$, or $\theta = \arcsin\sqrt{p}$. Tables VIII and IX give the θ-values for the obtained and experimental proportions respectively. The chi-square form applicable for this test of internal consistency is

$$\text{chi-square} = \frac{N}{821}\,\Sigma(\theta - \theta')^2$$

where N is the number of judgments per stimulus pair, θ is the arcsin $\sqrt{p}$ and θ' is the arcsin $\sqrt{p'}$. Mosteller (8) gives the degree of freedom for the above chi-square as

$$\text{df} = \frac{(n - 1)(n - 4)}{2}$$

where n is the number of stimuli. The sum of the squared differences between the observed and expected theta is 259.6927, and the number of pupils is forty-one; therefore

$$\text{Chi-square} = \frac{41(259.6927)}{821} \qquad \text{df} = 5$$

$$\text{Chi-square} = 12.969$$

Since the chi-square is not significant at the one percent level one could assume internal consistency between the observed and expected proportions for the preference of the six kindergarten activities as judged by the forty-one children in Class A. The scale-values for these activities are the $R_j$ values given in Table V, namely: Regular Stories .000, Music .017, Playing House .259, Working Puzzles .266, Easel Painting .524, and Causal Stories .1.053. The zero point is arbitrarily assigned.

Analysis of the preference of the six kindergarten activities for Class B, using Thurstone's Case III assumptions yielded slightly different results as obtained for Class A. Table II gives the proportion of time each activity was preferred over its respective pair. The transformation of these proportions into $R_j - R_k$ values where $R_j - R_k = z_{jk}\sqrt{s_j^2 + s_k^2}$ and the resultant scale-values are presented in Table X. Testing of the obtained

TABLE V

SCALE SEPARATIONS BETWEEN PAIRS OF SIX KINDERGARTEN ACTIVITIES AS JUDGED IN TERMS OF PREFERENCE BY THE FORTY-ONE KINDERGARTEN PUPILS IN CLASS A, ASSUMING CASE III (MATRIX $Z_{3A}$)

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories | Sum |
|---|---|---|---|---|---|---|---|
| Regular Stories | .000 | .229 | .129 | .116 | .479 | 1.166 | 2.119 |
| Music | - .229 | .000 | .822 | .272 | .198 | .956 | 2.019 |
| Playing House | - .129 | - .822 | .000 | .522 | .268 | .726 | .565 |
| Working Puzzles | - .116 | - .272 | - .252 | .000 | .806 | .625 | .521 |
| Easel Painting | - .479 | - .198 | - .268 | - .806 | .000 | .726 | -1.025 |
| Causal Stories | -1.166 | - .956 | - .726 | - .625 | - .726 | .000 | -4.199 |
| Sum $z_{jk}$ | -2.119 | -2.019 | - .565 | - .521 | 1.025 | 4.199 | .000 |
| Mean $z_{jk}$ | - .353 | - .336 | - .094 | - .087 | .171 | .700 | .001 |
| $R_j$ | .000* | .017 | .259 | .266 | .524 | 1.053 | 2.118 |

*Arbitrarily assigned zero point

TABLE VI

EXPECTED SCALE SEPARATIONS z' BETWEEN THE SIX KINDERGARTEN ACTIVITIES DERIVED FROM SCALE VALUES OBTAINED BY THE ASSUMPTIONS OF CASE III (MATRIX Z' A)

| Kindergarten Activities | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories |
|---|---|---|---|---|---|
| Regular Activities | .011 | .308 | .214 | .305 | 1.167 |
| Music | | .160 | .141 | .238 | .671 |
| Playing House | | | .006 | .152 | .844 |
| Working Puzzles | | | | .131 | .599 |
| Easel Painting | | | | | .299 |

TABLE VII

EXPECTED PROPORTIONS p' DERIVED THROUGH z' FROM SCALE VALUES FOR SIX KINDERGARTEN ACTIVITIES, AS JUDGED BY PUPILS IN CLASS A, UNDER THE ASSUMPTION OF CASE III (MATRIX P' A)

| Kindergarten Activities | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories |
|---|---|---|---|---|---|
| Regular Stories | .504 | .621 | .585 | .620 | .878 |
| Music | | .564 | .556 | .594 | .749 |
| Playing House | | | .502 | .560 | .801 |
| Working Puzzles | | | | .552 | .725 |
| Easel Painting | | | | | .618 |

TABLE VIII

OBTAINED PROPORTIONS FOR THE SIX KINDERGARTEN ACTIVITIES CONVERTED INTO ANGLES IN TERMS OF DEGREES (θ) BY MEANS OF THE FUNCTION $\theta = \text{ARCSIN}\sqrt{p}$ (MATRIX θ)

| Kindergarten Activities | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories |
|---|---|---|---|---|---|
| Regular Stories | 48.50 | 48.50 | 47.12 | 51.35 | 71.76 |
| Music | | 57.23 | 48.50 | 47.12 | 58.82 |
| Playing House | | | 54.27 | 48.50 | 62.03 |
| Working Puzzles | | | | 54.27 | 55.73 |
| Easel Painting | | | | | 54.27 |

θ values for Class A.

TABLE IX

EXPECTED PROPORTIONS FOR THE SIX KINDERGARTEN ACTIVITIES CONVERTED INTO ANGLES IN TERMS OF DEGREES (θ') BY MEANS OF THE FUNCTION $\theta' = \text{ARCSIN}\sqrt{p'}$ (MATRIX θ')

| Kindergarten Activities | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories |
|---|---|---|---|---|---|
| Regular Stories | 45.23 | 52.00 | 49.89 | 51.94 | 69.56 |
| Music | | 48.68 | 48.22 | 50.42 | 59.93 |
| Playing House | | | 45.11 | 48.45 | 63.51 |
| Working Puzzles | | | | 47.98 | 58.37 |
| Easel Painting | | | | | 51.83 |

θ' values for Class A.

## TABLE X

## SCALE SEPARATIONS BETWEEN PAIRS OF SIX KINDERGARTEN ACTIVITIES AS JUDGED IN TERMS OF PREFERENCE BY THE SIXTY-ONE KINDERGARTEN PUPILS IN CLASS B, ASSUMING CASE III (MATRIX $Z_{3B}$)

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories | Sum |
|---|---|---|---|---|---|---|---|
| Regular Stories | .000 | .222 | .648 | .416 | .966 | .740 | 2.993 |
| Music | - .222 | .000 | - .239 | - .152 | .644 | .632 | .663 |
| Playing House | - .648 | .239 | .000 | .610 | .147 | .645 | .993 |
| Working Puzzles | - .416 | .152 | - .610 | .000 | .665 | .318 | .109 |
| Easel Painting | - .966 | - .644 | - .147 | - .665 | .000 | - .171 | -2.593 |
| Causal Stories | - .741 | - .632 | - .645 | - .318 | .171 | .000 | -2.165 |
| Sum $z_{jk}$ | -2.993 | - .663 | - .993 | - .109 | 2.593 | 2.165 | .000 |
| Mean $z_{jk}$ | - .499 | - .110 | - .166 | - .018 | .432 | .361 | .000 |
| $R_j$ | .000* | .389 | .333 | .481 | .931 | .860 | 2.994 |

*Arbitrarily assigned zero point

## TABLE XI

## PROPORTION MATRIX FOR THE SIX KINDERGARTEN ACTIVITIES IN TERMS OF PREFERENCE AS JUDGED BY THE COMBINED CLASSES (MATRIX $P_C$)

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories |
|---|---|---|---|---|---|---|
| Regular Stories | .500 | .569 | .598 | .608 | .716 | .912 |
| Music | .431 | .500 | .559 | .500 | .608 | .716 |
| Playing House | .402 | .441 | .500 | .627 | .539 | .686 |
| Working Puzzles | .392 | .500 | .373 | .500 | .667 | .647 |
| Easel Painting | .284 | .392 | .461 | .333 | .500 | .529 |
| Causal Stories | .088 | .284 | .314 | .353 | .471 | .500 |
| $p_{j>k}$ | 2.097 | 2.686 | 2.805 | 2.921 | 3.501 | 3.990 |

## TABLE XII

## SCALE SEPARATIONS BETWEEN PAIRS OF SIX KINDERGARTEN ACTIVITIES AS JUDGED IN TERMS OF PREFERENCE BY THE 102 PUPILS OF THE COMBINED CLASSES, ASSUMING CASE III (MATRIX $Z_{3C}$)

| Kindergarten Activities | Regular Stories | Music | Playing House | Working Puzzles | Easel Painting | Causal Stories | Sum |
|---|---|---|---|---|---|---|---|
| Regular Stories | .000 | .259 | .327 | .289 | .900 | .702 | 2.567 |
| Music | -.259 | .000 | .281 | .000 | .571 | .849 | 1.442 |
| Playing House | -.327 | -.281 | .000 | .513 | .193 | .639 | .737 |
| Working Puzzles | -.289 | .000 | -.513 | .000 | .779 | .398 | .375 |
| Easel Painting | -.900 | -.571 | -.193 | -.779 | .000 | .115 | 2.328 |
| Causal Stories | -.792 | -.849 | -.639 | -.398 | -.115 | .000 | -2.793 |
| Total | -2.567 | -1.442 | -.737 | -.375 | 2.328 | 2.793 | .000 |
| Mean $z_j$ | -.428 | -.240 | -.123 | -.062 | .388 | .466 | .001 |
| $R_j$ | .000* | .188 | .305 | .366 | .816 | .894 | 2.569 |

*Arbitrarily assigned zero point

## TABLE XIII

## SCALE VALUES FOR THE SIX KINDERGARTEN ACTIVITIES COMPUTED BY THE METHOD OF PAIRED COMPARISONS DERIVED BY THE ASSUMPTIONS OF THURSTONE'S CASE III

| Kindergarten Activities | Class A | Class B | Combined Class A and B |
|---|---|---|---|
| Causal Stories | 1.053 | .860 | .894 |
| Easel Paintings | .524 | .931 | .816 |
| Working Puzzles | .266 | .481 | .366 |
| Playing House | .259 | .333 | .305 |
| Music | .017 | .389 | .188 |
| Regular Stories | .000* | .000* | .000* |

*Arbitrarily assigned zero point

and expected proportions by Mosteller's method showed internal consistency of the data at the one percent level, the level chosen to reject the hypothesis. The scale-values computed for the six kindergarten activities as derived by means of paired comparisons and judged by the sixty-one pupils in Class B were Regular Stories .000, Playing House .333, Music .389, Working Puzzles .481, Causal Stories .860, and Easel Painting .931. Two pairs of activities reversed position when compared with scale-values for the kindergarten activities as judged by the children in Class A, namely, Music and Playing House, and Easel Painting and the Causal Stories.

As another part of the analysis of the investigation, the data for Class A and Class B were pooled and scale-values for the combined group were calculated. Table XI gives the obtained proportions for the combined group.

Thurstone's Case III assumptions were applied to the proportions obtained by pooling the data. The $R_j$ - $R_k$ values under these assumptions are presented in Table XII plus the scale-values for each of the kindergarten activities.

The chi-square test for internal consistency was not significant at the one percent level. The scale-values thus obtained by Thurstone's Case III assumptions for the combined classes were Regular Stories .000, Music .188, Playing House .305, Working Puzzles .366, Easel Painting .816, and Causal Stories .894. The zero point is an arbitrarily assigned value.

## Conclusions and Summary

The purpose of this experiment was to find out how well kindergarten children liked to listen to the causally oriented stories specially written for their age. Subjects were forty-one kindergarten children in Class A and sixty-one in Class B. Thurstone's Case III assumptions were used. After comparing this activity with five other kindergarten activities by means of paired comparisons, the results showed that listening to the prepared causal stories was preferred by Class A and the combined group over the other five kindergarten activities, namely, having Music, Working Puzzles, Painting at the Easel, Playing House, or Having Regular Kindergarten Stories read to them. The significance of internal consistency was tested by using the one percent level of rejection. A similar analysis of the paired comparison data for Class B showed a reversal in the preference of Causal Stories and Easel Painting. Easel Painting rated first and causal narratives, second.

Thus, using the materials designed to help the child develop a causal orientation toward his social environment ranked well with other activities included in this investigation.

Table XIII is a summary of the scale values assigned to the six kindergarten activities as judged by pupils in Class A, Class B, and the Combined Classes A and B, using Thurstone's Case III assumptions for scaling the paired comparison data.

## REFERENCES

1. Burros, R. H. "The Application of the Method of Paired Comparisons to the Study of Reaction Potentials," Psychological Review LVIII (1951), pp. 60-66.
2. Guilford, J. P. "The Method of Paired Comparisons as a Psychometric Method," Psychological Review, XXXV (1928), pp. 494-506.
3. Guilford, J. P. "Some Empirical Tests of the Method of Paired Comparisons," Journal of General Psychology, V (1931), pp. 64-76.
4. Guilford, J. P. Psychometric Methods, Second Edition (New York: McGraw Hill Book Co., 1954), pp. 154-77.
5. Levitt, E. E., and Ojemann, R. H. "The Aims of Preventive Psychiatry and 'Causality' as a Personality Pattern," Journal of Psychology, XXXVI (1953), pp. 393-400.
6. Mosteller, F. "Remarks on the Method of Paired Comparisons: I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations," Psychometrika, XVI (1951), pp. 3-9.
7. Mosteller, F. "Remarks on the Method of Paired Comparisons: II. The Effect of an Aberrant Standard Deviation When Equal Standard Deviations and Equal Correlations are Assumed," Psychometrika, XVI (1951), pp. 203-6.
8. Mosteller, F. "Remarks on the Method of Paired Comparisons: III. A Test of Significance for Paired Comparisons When Equal Standard Deviations and Equal Correlations are Assumed," Psychometrika, XVI (1951), pp. 207-218.
9. Ojemann, R. H., and others. "The Effects of a 'Causal' Teacher-Training Program and Certain Curricular Changes on Grade School Children," Journal of Experimental Psychology, XXXVI (1953), pp. 393-400.
10. Ross, R. T. "Optimum Orders for the Presentation of Pairs in the Method of Paired Comparisons," Journal of Educational Psychology, XXV (1934), pp. 375-82.
11. Ross, R. T. "Optimal Orders in the Method of Paired Comparisons," Journal of Experimental Psychology, XXV (1939), pp. 414-24.
12. Thurstone, L. L. "Psychophysical Analysis," American Journal of Psychology, XXXVIII (1927), pp. 368-89.
13. Thurstone, L. L. "A Law of Comparative

Judgment," Psychological Review, XXXIV (1927), pp. 273-86.

14. Thurstone, L. L. "The Method of Paired Comparisons for Social Values," Journal of Abnormal and Social Psychology, XXI (1927), pp. 384-400.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# ANALYSIS OF MMPI DATA FOR CLASSIFICATION PURPOSES BY MULTIVARIATE STATISTICAL TECHNIQUES*

PETER P. REMPEL
State University of Iowa

THE PURPOSE of this paper is to present the steps involved in the application of multivariate analysis to the assignment—with a minimum error of classification—of N people to q criterion groups on the basis of all information provided by k intercorrelated measures.

The study itself stems from the writer's attempt to identify as potential delinquents or nondelinquents (q = 2), a population of ninth-grade high-school boys (N = 1802) to whom the MMPI had been administered by Hathaway (3) in 1947-1948, and for whom police and/or court records were available through 1952. Severity of misdemeanor made possible a classification of delinquency into eight levels. Because of inter-group similarity of MMPI profiles between nondelinquents and delinquents at levels one and two, the latter (N = 363) have been omitted from this study (see Table I).

Variables used in one part of the study, referred to as the main study, consisted of the scores on the F, Hs, Sc, Mf, and Pd scales of the MMPI (k = 5); in another part, the school study, they consisted of the Ma, Sc, Mf, and Pd scales of the MMPI; honor point ratio, and numbers of days absent in Grade VIII (k = 6).

In the main study the MMPI profiles of 351 boys who had achieved delinquency status by 1952 were divided by random procedures into two groups: Sample A (N = 175) which was used in the development of the classification formula, and Sample B (N = 176), set aside for cross-validation purposes. Two samples of 175 profiles each were similarly selected from the nondelinquent set of MMPI profiles. In the school study both delinquent and nondelinquent groups were divided into Samples A' and B' and four sub-groups (N = 91) were thus similarly obtained.

## Tests of Assumption

Numerical values obtained for the statistics $\beta_1^{\frac{1}{2}}$ and $\beta_2$—the former used as a measure of skewness, the latter as a measure of kurtosis—indicated that roughly the distributions of only half of the variables tested in either study (in Samples A and A') satisfied the assumption of normality. Somewhat more satisfying results were obtained from tests of homogeneity of variances and covariances. Thus approximately 30 percent of the tests indicated lack of homogeneity of variances at a statistically significant level, and 20 percent indicated lack of homogeneity of covariances. The effect of nonnormality upon the distribution of $D^2$ and the LDF statistic is not known. A test of the assumption of linearity of regression (4:240-245) was omitted because the number of variables involved made the amount of labor required too excessive. Inspection of correlation charts of selected pairs of variables made it appear unlikely that anything other than linearity existed. Because cross-validation has been included, the application of multivariate analysis was thought to be justified even though not all of the assumptions have been fully met. But in drawing conclusions from this study, one must recognize the fact that not all of the assumptions for the application of multivariate procedures were fulfilled.

## Discussion of Statistical Techniques

The generalized distance function contributes both to the problem of testing significance and of estimation. The five (or six) variables used in the main and school studies respectively, may be considered as defining a space of five (or six) dimensions so that any individual with assigned scores can be represented by a point in such a space. The individuals belonging to any group will then be represented by a cluster of points round another point which represents the mean values of these scores. If the mean values of the two groups are close together, then the clusters of points corresponding to them will overlap to some extent. Rao (6) has shown that the degree of overlap between any two

---

*Submitted to the Graduate School of the University of Minnesota in partial fulfillment of the requirements for the Ph. D. degree.

TABLE I

DISTRIBUTION OF VALID MMPI PROFILES BY DELINQUENTS, LEVEL OF DELINQUENCY, AND NON-DELINQUENTS

| Classification | Number | Percent (Based on N = 1802) | Percent (Based on N = 714) |
|---|---|---|---|
| All Delinquents | 714 | 39.62 | |
| Levels 1 and 2 | 363 | 20.15 | 50.84 |
| Level 3 | 231 | 12.82 | 32.35 |
| Level 4 | 47 | 2.61 | 6.58 |
| Level 5 | 31 | 1.72 | 4.34 |
| Level 6 | 22 | 1.22 | 3.08 |
| Level 7 | 12 | 0.67 | 1.68 |
| Level 8 | 8 | 0.44 | 1.12 |
| Total Level 3-8 | 351 | 19.47 | 49.16 |
| Non-delinquents | 1088 | 60.37 | ..... |
| Total (All delinquents and non-delinquents) | 1802 | 100.00 | ..... |

TABLE II

A SUMMARY COMPARISON OF CLASSIFICATION EFFICIENCY FOR THE LDF PROCEDURE WITH CHANCE EXPECTATION FOR SAMPLES A AND B IN THE MAIN STUDY BY DELINQUENT AND NON-DELINQUENT GROUPS WITHIN THE SAMPLE

| Samples Used | | Correct Classification by Formula | | Correct Classification by Chance | | Percentage Improvement Over Chance |
|---|---|---|---|---|---|---|
| | | N | % | N | % | |
| Sample A - Delinquents | N = 175 | 120 | 68.6 | 87.5 | 50.0 | 18.6** |
| Sample A - Non-delinquents | N = 175 | 115 | 65.7 | 87.5 | 50.0 | 15.7** |
| Total Sample A | N = 350 | 235 | 67.1 | 175.0 | 50.0 | 17.1** |
| Sample B - Delinquents | N = 176 | 124 | 70.5 | 88.0 | 50.0 | 20.5** |
| Sample B - Non-delinquents | N = 175 | 103 | 58.9 | 87.5 | 50.0 | 8.9* |
| Total Sample B | N = 351 | 227 | 64.7 | 175.5 | 50.0 | 14.7** |
| All Groups Total | N = 701 | 462 | 65.9 | 350.5 | 50.0 | 15.9** |

* Significant at the 5 percent level
**Significant at the 1 percent level

such clusters can be measured by the quantity

$$D^2 = \sum_{i=1}^{i=5(6)} \sum_{j=1}^{j=5(6)} a^{ij} d_i d_j$$

where $d_1 d_2; \ldots, d_5$ (or $d_6$) are the differences between the mean values of the measurements in the two groups, and $a^{ij}$ are the elements of the matrix reciprocal to the within group dispersion matrix $(a_{ij})$. The higher the value of $D^2$, the smaller the amount of overlap, so that this measure may be regarded as the "distance" between the groups.

The calculation of $D^2$ is given in the basic form as

$$D^2 = \Sigma_i \Sigma_j S^{ij} d_i d_j \tag{1}$$

$S^{ij}$ = elements of the matrix inverse to $S_{ij}$, the dispersion or within groups variance-covariance matrix.

$d_i = \overline{X}_i - \overline{X}_i'$

$d_j = \overline{X}_j - \overline{X}_j'$

where $\overline{X}_i$ represents the sample mean of the ith variable and $\overline{X}_i'$ is the sample mean of a second category of the ith variable.

The "best" (unbiased) estimate of the sample variance used in the matrix for the ith test is

$$\sigma_i{}^2 = \frac{1}{n}\left[\sum_{k=1}^{N_x} X_{ik}^2 - N_x \overline{X}_i{}^2 + \sum_{k=1}^{N_y} Y_{ik}^2 - N\overline{Y}_i^2\right] \tag{2}$$

and the unbiased estimate of the covariance becomes

$$\rho_{ij} = \rho_{ji} = \frac{1}{n}\left[\sum_{k=1}^{N_x} X_{ik} X_{jk} - N_x \overline{X}_i \overline{X}_j + \sum_{k=1}^{N_y} Y_{ik} Y_{jk} - N_y \overline{X}_i \overline{X}_j\right] \tag{3}$$

where

$X_i$ $(i = 1, 2, \ldots, r)$ = the random variable defined by the ith test (or scale) in the first group,

$Y_i$ $(i = 1, 2, \ldots, r)$ = the random variable defined by the ith test (or scale) in the second group,

$X_{ik}$ = the score of the kth individual in the first group on the ith test.

$Y_{ik}$ = the score of the kth individual in the second group on the ith test,

$\overline{X}_i$ = the sample mean of the ith test in the first group, and

$\overline{Y}_i$ = the sample mean of the ith test in the second group.

Rao (9) met the problem of eliminating the correlation between variables by suggesting a method of transforming the original set of data to an orthogonal set of variates. The most efficient method of accomplishing this purpose is the application of pivotal condensation to the calculation of $D^2$, and is the method used in this study. A systematic and straight-forward illustration of the use of Rao's pivotal condensation technique is found in a paper by Collier (2).

The application of pivotal condensation to the problem of transforming correlated variates to a new orthogonal set reduces the covariance elements in the $S_{ij}$ variance-covariance matrix to zero, thus eliminating the element of correlation from matrix calculations. The general distance function may then be expressed as

$$D^2 = \frac{\Sigma {*} d_i^2}{{*}S_{ii}} , \tag{4}$$

in which $*S_{ii}$, the variance of the transformed ith variable is found easily in the process of pivotal condensation, and each $\frac{*d_i{}^2}{*S_{ii}}$ represents the contribution of the ith variable to the total $D^2$.

### Significance of Added Variables

Rao (7, 8, 9) has considered the following approximate test of significance which can be used for testing both the significance of the total $D^2$ value as well as of the addition to $D^2$ of each variable separately:

$$T = \frac{(f-p-q+1)}{q}\left[\frac{n_1 n_2 (D^2{}_{p+q} - D^2{}_p)}{f(n_1+n_2)+n_1 n_2 D^2{}_p}\right] \tag{5}$$

where f = the degrees of freedom of the dispersion matrix of variance and covariances between the ith and jth characters

p = number of variables in $D_p^2$
q = number of variables added to form $D_{p+q}^2$
$n_1$ = size of sample or groups from the first population considered
$n_2$ = size of sample from the second population

T has the F-distribution with q and f-p-q+1 degrees of freedom. Use of this value enables the research worker to determine the effect of adding variables to the matrix.

## The Linear Discriminant Function (LDF)

Let $X_{ijk}$ be a typical observation where i = 1, 2 stands for the sample, j for the trait measured, and k for the individual. Then the value Y, would be determined by the function

$$Y_{ik} = \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_p X_p \tag{6}$$

and

$$\overline{Y}_{ik} = \frac{i}{n_i} \Sigma_i Y_{ik} \tag{7}$$

where $\Sigma_i$ denotes summation over the ith sample (i = 1, 2). The problem is to maximize the ratio

$$\frac{\overline{Y}_2 - \overline{Y}_1}{\sqrt{\Sigma_1 (Y_{1k} - Y_1)^2 + \Sigma_2 (Y_{2k} - Y_2)^2}} \tag{8}$$

which may be written as

$$\frac{\left( \sum_{j=1}^{p} \lambda_j d_j \right)^2}{\sum_{j=1}^{p} \sum_{m=1}^{p} \lambda_j \lambda_m S_{jm}} \tag{9}$$

where $d_j = \frac{1}{n_2} \Sigma_2 X_{2jk} - \frac{1}{n_1} \Sigma_1 X_{1jk}$ for the jth quantity (j = 1, 2, . . . , p): and $S_{jm}$ is the pooled sum of products of deviations from the sample traits j and m. Maximization of the above ratio requires essentially setting the derivative of the ratio equal to zero, and we obtain

$$\frac{\sum_{m=1}^{p} \sum_{j=1}^{p} \lambda_j \lambda_m S_{jm}}{\sum_{j=1}^{p} \lambda_j d_j} \cdot dr = \sum_{j=1}^{p} \lambda_j \lambda_{jr}, \quad (r = 1, 2, \dots, p) \tag{10}$$

Since the first factor in the above equation will be the same for all equations it can be ignored. Therefore, the coefficients required are obtained through the solutions of the normal equations

$$\left. \begin{array}{l} S_{11}\lambda_1 + S_{12}\lambda_2 \dots + S_{1p}\lambda_p = d_1 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ S_{p1}\lambda_1 + S_{p2}\lambda_2 \dots + S_{pp}\lambda_p = d_p \end{array} \right] \tag{11}$$

The diagonal cells in the above matrix represent variances, while all other cells consist of the appropriate covariances. $D^2$ is represented by the expression

$$D^2 = \sum_{i=1}^{p} \lambda_i d_i \tag{12}$$

and the discriminant function by

$$\text{LDF} = \sum_{i=1}^{p} \lambda_i X_i \tag{13}$$

## Classification Procedures

The application of Fisher's discriminant function to the problem of classification into one or two known multivariate normal populations is discussed by Anderson (1). Suppose that an individual I with a set of measurements $X_1, \dots, X_q$ has been randomly drawn from one of two populations $\Pi_1$ and $\Pi_2$, with probability distributions of the measurements $f_1(X_1, \dots, X_q)$, respectively. Suppose further that $X_1, \dots, X_q$ has a joint normal distribution with means in $\Pi_1$ of $EX_i = u_i$ (1), and in $\Pi_2$ of $EX_i = u_i$ (2). Let the common set of variances and covariances be

$$\sigma_1^2, \dots, \sigma_{p_i}^2\ \rho_{12},\ \rho_{13}, \dots \rho_{p-1},\ p.$$

It can then be shown that the best regions of classification are of the general form

$$R_1 : \sum_{i=1}^{p} \lambda_i X_i > \sum_{i=1}^{p} \lambda_i \tfrac{1}{2} (u_i^{(1)} + u_i^{(2)}) + \log k \tag{14}$$

$$R_2 : \sum_{i=1}^{p} \lambda_i X_i < \sum_{i=1}^{p} \lambda_i \tfrac{1}{2} (u_i^{(1)} + u_i^{(2)}) + \log k \tag{15}$$

where $\lambda_1, \dots, \lambda_p$ is the solution of

$$\sum_{j=1}^{p} \sigma_i \sigma_j \rho_{ij} \lambda_j = u_i^{(1)} - u_i^{(2)} \tag{16}$$

and $\log k = \log \dfrac{L(1 \mid 2) p_2}{L(2 \mid 1) p_1}$

where $L(2 \mid 1)$ is defined as the loss incurred in classifying individual I into $\Pi_2$ when in fact he belongs to $\Pi_1$, and $L(1 \mid 2)$ as the loss incurred in classifying individual I into $\Pi_1$ when in fact he belong to $\Pi_2$.

The first term of Formulas 14 and 15 is the same as that to which the LDF simplifies (see Formula 13) if the population expectations and covariance matrix are known. Therefore if k = 1, i. e., if the probability of belonging to one group is equal to coming from the other, then log k = 0, and the procedure is reduced to a comparison of the LDF of individual observations with the LDF of the averages of the respective means. If a priori probabilities are not specified. log k is found so that the risk when I is from $\Pi_1$ is equal to the risk when I is from $\Pi_2$. For this purpose let

$$\sum_i^q \lambda_i X_i - \tfrac{1}{2}\sum_i^q \lambda_i (u_i^{(1)} + u_i^{(2)}) = U(X_1, \ldots, X_q) \qquad (17)$$

Since a linear function of normally distributed variables is again normally distributed, the distribution of U is likewise normal, with the mean and variances found as follows:

$$E(U) = \sum_i^q \lambda_i \, E(X_i) - \tfrac{1}{2}\sum_i^q \lambda_i (u_i^{(1)} + u_i^{(2)}) \qquad (18)$$

$$\text{and} \quad \sigma_U^2 = \sum_i^q \lambda_i \, (u_i^{(1)} - u_i^{(2)}) \qquad (19)$$

$$\text{Let} \quad \sigma_U^2 = a \qquad (20)$$

Then if I is from $\Pi_1$, $E(X_i) = u_i^{(1)}$ and

$$E(U) = \tfrac{1}{2}\sum_i^q \lambda_i \, (u_i^{(1)} - u_i^{(2)}) = \tfrac{1}{2}a \qquad (21)$$

where a is the "distance" between $\Pi_1$ and $\Pi_2$ or the variance of U. If I is from $\Pi_2$, $E(X_i) = u_i^{(2)}$, and

$$E(U) = \sum_i^q \lambda_i \, (u_i^{(1)} - u_i^{(2)}) = -\tfrac{1}{2}a \qquad (22)$$

The probability of misclassifying an individual belonging to $\Pi_1$ is then determined by referring the following relation to a normal distribution table:

$$P\left( Z < \frac{\log_e k - \frac{1}{2}a}{\sqrt{a}} \right) \qquad (23)$$

The probability of misclassifying an individual belonging to $\Pi_2$ is determined by referring the following relation to a normal distribution table:

$$P\left( Z > \frac{\log_e k - \frac{1}{2}a}{\sqrt{a}} \right) \qquad (24)$$

In both cases $k = \Pi_2 \;\; \Pi_1$. If $\Pi_2 = \Pi_1$ then $k = 1$, and $\log k = 0$. Then the formula for calculating the theoretical probability of misclassification is reduced to

$$P\left( Z > \frac{o - (-\frac{1}{2}a)}{\sqrt{a}} \right) \qquad (25)$$

The overall probability of misclassification is then the sum of products of the population proportions in each group and the respective misclassification probabilities.

## Test of Effectiveness of Linear Discriminant Functions

Lubin (5) has given us the tests of the effectiveness of the LDF in classifying individuals correctly. In a two-contingency table let the columns $A_1$ and $A_2$ represent the observed categories while the rows $B_1$ and $B_2$ represent the predicted categories. If perfect discrimination is achieved, only the diagonal cells will be filled. For pure chance discrimination

$$n_i = \frac{A_i B_j}{N} \qquad (26)$$

where

$n_{ij}$ = frequency of the ith row and the jth column
$A_i$ = marginal total of the ith row
$B_j$ = marginal total of the jth row
$N$ = total number of individuals involved.

In case the value of chi-square computed on the diagonal values is found to be significant (and a chi-square test is a necessary but not sufficient test to demonstrate classification beyond chance), then the one-tailed t-test for large values of N, may be used to demonstrate the effectiveness of the LDF:

$$t = \frac{(o - e)\sqrt{N}}{\sqrt{e(N - e)}} \qquad (27)$$

where o denotes the number of individuals classified correctly, and e the number of correct classifications expected by chance.

## Computational Procedures

In the main study the variance-covariance matrix representing the equation

$$\sum_{j=1}^{5} \sigma_i \sigma_j \rho_{ij} \lambda_j = u_i^{(1)} - u_i^{(2)}, \quad i = 1, \ldots, 5$$

was found to be as shown at top of next page.

This set of simultaneous linear equations for Sample A was solved by the method of pivotal condensation. This method reduces the original set of variables to a new orthogonal set of uncorrelated variables, and makes possible a test of significance of the contribution of each variable to $D^2$. The numerical values of the beta weights and the individual contribution of variables to $D^2$ are as shown in the middle of the next page.

The value of $D^2$ computed from the product of these weights and the difference in means for each variable was found to be .9173 (i. e., the same that

|  | F | Hs | Sc | Mf | Pd | $u_i^{(1)} - u_i^{(2)}$ |
|---|---|---|---|---|---|---|
| F | $11.9238\lambda_1 +$ | $1.9755\lambda_2 +$ | $17.1353\lambda_3 +$ | $3.5074\lambda_4 +$ | $8.4253\lambda_5 =$ | 1.5029 |
| Hs | $1.9755\lambda_1 +$ | $77.1411\lambda_2 +$ | $38.6356\lambda_3 +$ | $19.5887\lambda_4 +$ | $30.8370\lambda_5 =$ | 2.2743 |
| Sc | $17.1353\lambda_1 +$ | $38.6256\lambda_2 +$ | $104.1166\lambda_3 +$ | $28.5664\lambda_4 +$ | $50.1619\lambda_5 =$ | 4.9714 |
| Mf | $3.5074\lambda_1 +$ | $19.5887\lambda_2 +$ | $28.5664\lambda_3 +$ | $72.7225\lambda_4 +$ | $16.9858\lambda_5 =$ | -3.5029 |
| Pd | $8.4253\lambda_1 +$ | $30.8370\lambda_2 +$ | $50.1619\lambda_3 +$ | $16.9858\lambda_4 +$ | $99.6932\lambda_5 =$ | 6.6629 |

| Scale | Beta Weights | Contribution of Variables to $D^2$ |
|---|---|---|
| F | $\lambda_1 = .0695$ | .1894** |
| Hs | $\lambda_2 = .0116$ | .0534* |
| Sc | $\lambda_3 = .0256$ | .0556* |
| Mf | $\lambda_4 = -.0782$ | .3750** |
| Pd | $\lambda_5 = .0578$ | .2439** |
|  |  | $D^2$ = .9173 |

was obtained by adding the contributions of variables to $D^2$ above).

The first part of the expression

$$U = \sum_{i=1}^{p} \lambda_i X_i - \sum_{i=1}^{p} \lambda_i \tfrac{1}{2}(u_i^{(1)} + u_i^{(2)})$$

is obviously the LDF, and for any set of scores in Sample A, its numerical value will be written in terms of the beta weights for $D^2$ as follows:

$$LDF = \sum_{i=1}^{p} \lambda_i X_i = .0695X_1 + .0116X_2 + .0256X_3 - .0782X_4 + .0578X_5$$

The second part of the expression represents the middle point between the sums of the weighted means of the two criterion groups and is therefore a constant quantity, c. For Sample A the value of c is the sum of the products of the beta weights and the sum of means, or

$$c = \tfrac{1}{2}\left[(.0695)(12.309) + (.0116)(103.4514) + (.0256)(119.0628) + (-.0782)(103.6629) + (.0578)(121.6914)\right] = 2.0169$$

Also, since

$$\sigma_U^2 = \sum_{i}^{q} \lambda_i (u_i^{(1)} - u_i^{(2)}) = a,$$

and

$$\sum_{i}^{q} \lambda_i (u_i^{(1)} - u_i^{(2)})$$

is identical with the expression $\sum_{i=1}^{q} \lambda_i d_i$ it follows that $\sigma_U^2 = D^2$, or $a = 0.9173$, and $a^{\frac{1}{2}} = 0.9578$. For the delinquent sample therefore, the mean of U turns out to be $\frac{1}{2}a$; for the non-delinquent sample the mean is $-\frac{1}{2}a$. The numerical value of these two statistics was found to be a = .4586 and -a = -.4586.

## Solution of the Minimum Probability of Misclassification

By re-working Formulas 23 and 24 above, through the use of Formula 25, the probability of misclassifying an individual from either population is found to be .3161. Thus

## TABLE III

### CLASSIFICATION OF DELINQUENTS BY LEVEL OF DELINQUENCY USING LDF PROCEDURES FOR TWO RANDOMLY CHOSEN SAMPLES OF THE MAIN STUDY: SAMPLE A ON WHICH PREDICTION FORMULAE WERE DEVELOPED AND SAMPLE B ON WHICH THE FORMULAE WERE APPLIED THROUGH CROSS-VALIDATION

| | | Percentage of Correct Classification | | | | | |
|---|---|---|---|---|---|---|---|
| | | Sample A (N = 175) | | Sample B (N = 176) | | Total (N = 351) | |
| Level* | Total | N | % | N | % | N | % |
| 3 | 231 | 82 | 70.7 | 78 | 67.8 | 160 | 69.3 |
| 4 | 47 | 14. | 60.9 | 18 | 75.0 | 32 | 68.1 |
| 5 | 31 | 7 | 46.7 | 12 | 75.0 | 19 | 61.3 |
| 6 | 22 | 9 | 81.8 | 8 | 73.0 | 17 | 77.3 |
| 7 | 12 | 4 | 66.7 | 5 | 83.3 | 9 | 75.0 |
| 8 | 8 | 4 | 100.0 | 3 | 75.0 | 7 | 87.5 |
| Total | 351 | 120 | 68.6 | 124 | 70.5 | 244 | 69.5 |

*In addition to the 351 delinquents in levels 3 to 8 above which had been randomly assigned to Samples A and B, 363 pupils had been classified as delinquents of levels 1 and 2. Application of the same procedures identified 75 or 42.3 percent of the level 1 group as delinquent and 111 or 60.0 percent of the level 2 group, an overall correct classification of 51.2 percent, approximately the same as chance expectation.

## TABLE IV

## A SUMMARY COMPARISON OF CLASSIFICATION EFFICIENCY FOR THE LDF PROCEDURES WITH CHANCE EXPECTATIONS FOR SAMPLES A' AND B' IN THE SCHOOL STUDY BY DELINQUENT AND NON-DELINQUENT GROUPS WITHIN THE SAMPLE

| Samples Used | | Correct Classification by Formula N | % | Correct Classification by Chance N | % | Percentage Improvement Over Chance |
|---|---|---|---|---|---|---|
| Sample A' - Delinquents | N = 91 | 60 | 65.9 | 45.5 | 50.0 | 15.9* |
| Sample A' - Non-delinquents | N = 91 | 72 | 79.1 | 45.5 | 50.0 | 29.1* |
| Total Sample A' | N = 182 | 132 | 72.5 | 91.0 | 50.0 | 22.5* |
| Sample B' - Delinquents | N = 91 | 63 | 69.2 | 45.5 | 50.0 | 19.2* |
| Sample B' - Non-delinquents | N = 91 | 63 | 69.2 | 45.5 | 50.0 | 19.2* |
| Total Sample B' | N = 182 | 126 | 69.2 | 91.0 | 50.0 | 19.2* |
| All Groups Total | N = 364 | 258 | 70.9 | 182.0 | 50.0 | 20.9* |

*Significant at the 1 percent level

## TABLE V

## CLASSIFICATION OF DELINQUENTS BY LEVEL OF DELINQUENCY USING LDF PROCEDURES FOR TWO SAMPLES OF THE SCHOOL STUDY: SAMPLE A' ON WHICH PREDICTION FORMULAE WERE DEVELOPED AND SAMPLE B' ON WHICH THE FORMULAE WERE APPLIED THROUGH CROSS-VALIDATION

| Level | Total | Sample A' (N = 91) N | % | Sample B' (N = 91) N | % | Total (N = 182) N | % |
|---|---|---|---|---|---|---|---|
| 3 | 106 | 38 | 71.70 | 31 | 59.62 | 69 | 65.71 |
| 4 | 30 | 9 | 60.00 | 12 | 75.00 | 21 | 67.74 |
| 5 | 21 | 9 | 75.00 | 6 | 66.67 | 15 | 71.43 |
| 6 | 13 | 5 | 71.43 | 5 | 83.33 | 10 | 76.97 |
| 7 | 7 | 1 | 50.00 | 3 | 60.00 | 4 | 57.14 |
| 8 | 5 | 1 | 50.00 | 3 | 100.00 | 4 | 80.00 |
| Total | 182 | 63 | 69.23 | 60 | 65.93 | 123 | 67.58 |

$$p_1 = p_2 = .50$$
$$k = p_2/p_1 = 1.00$$
$$\log_e k = 0.00$$

and $\Pi_1 = \Pi_2 = p\left(Z > \frac{0.0 - (-.4586)}{.9578}\right)$ or .3161

which is equal to 31.61 percent.

This means, in effect, that the theoretical probability of correct classification for both delinquents and non-delinquents is 68.39 percent. This overall percentage is essentially the same as that achieved by chance when the ratio of non-delinquents to delinquent in the population is known to be 80.5 to 19.5, as it is in this study. However, classification of delinquents by the LDF is more than three times as effective as chance procedures, the formula identifying 68.4 percent of the delinquents as against 19.5 percent identified by chance.

In actual practice, however, the population distribution is not known. In such a case, pure chance procedures make possible the identification of a maximum of 50.00 percent of each of the two populations.

## Empirical Cross-Validation of Samples A and B

The actual effectiveness of the linear discriminant function was tested empirically by applying the formulas developed on Sample A to the delinquent and non-delinquent sub-samples in both Samples A and B. A study of cross-validation results as presented in Table II indicates that the overall percentage of correct classifications, namely 65.9, fell somewhat short of the predicted percentage of 68.4. The fact that the formula correctly identified only 58.9 percent of the non-delinquents in Sample B accounts for most of the shrinkage. However, 68.6 percent of delinquents in Sample A, and 70.5 percent in Sample B were correctly identified.

Classification results were also broken down in terms of delinquency level, as presented in Table III. The tabulation shows generally increasing proportions of correct classification with level of delinquency, level 5 being the only exception.

## Results of the School Study

The set of linear equations composing the variance-covariance matrix for the six variables used in the school study is presented at the top of the next page.

The following is a list of the essential statistics derived:

$$\text{LDF} = .0219X_1 + .0282X_2 - .0430X_3 + .0398X_4 - .5247X_5 + .0564X_6$$
$$c = 2.7955$$

$$a = 1.2387; \sqrt{a} = 1.1129; \tfrac{1}{2}a = .6194$$

Using Formula 25 the probability of misclassifying an individual from either population is found to be equal to

$$\Pi_1 = \Pi_2 = p\ (Z > \frac{0.0 - (-.6194)}{1.1129}) =$$

$$p\ (Z > .5566) = .2906$$

or approximately 29.1 percent, or a probability of correct classification for both delinquents and non-delinquents of 70.9 percent.

Results of cross-validation procedures (see Table IV) indicate that classification by the LDF permitted the correct identification of 65.9 percent of the delinquents in Sample A' and of 69.2 percent in Sample B'. Corresponding figures for non-delinquents are 79.1 percent and 69.2 percent for Samples A' and B' respectively. In each case improvement over chance is significant at the one percent level.

Although the two studies are not strictly comparable, the addition of school record data raised the overall predictive efficiency of the LDF from 65.9 percent in the main study to 70.9 percent in the school study, a difference significant at the 5 percent level.

Classification results in terms of the level of delinquency are presented by means of Table V. Again, an increase in the proportion of correct classifications corresponds closely with increase in level of delinquency, a marked drop in predictive efficiency at level 7 being the only exception in the general trend.

## Summary

This study has demonstrated the usefulness of applying multivariate statistical techniques to the analysis of MMPI scale scores alone, or in combination with school data, for the purpose of classifying ninth-grade boys as potential delinquents or non-delinquents. The techniques employed proved to be effective to the extent that 62.3 percent of the non-delinquents and 69.5 percent of the delinquents were correctly identified by the use of Multiphasic data alone, while a combination of Multiphasic and school record data made possible the correct identification of 74.2 percent of non-delinquent boys and 67.5 percent of delinquent-prone boys.

## REFERENCES

1. Anderson, T. W. "Classification by Multivariate Analysis," *Psychometrika*, XVI (1951), pp. 31-50.
2. Collier, R. O. "Some Applications of the Method of Pivotal Condensation in Statistical Anal-

| | Ma | Sc | Mf | Pd | HPR | Days | |
|---|---|---|---|---|---|---|---|
| Ma | $125.8646\lambda_1 +$ | $42.2858\lambda_2 +$ | $8.5737\lambda_3 +$ | $26.3314\lambda_4 -$ | $.8156\lambda_5 -$ | $6.9509\lambda_6 =$ | $5.4505$ |
| Sc | $42.2858\lambda_1 +$ | $107.1731\lambda_2 +$ | $19.2574\lambda_3 +$ | $37.0621\lambda_4 -$ | $.6524\lambda_5 +$ | $4.7766\lambda_6 =$ | $5.2088$ |
| Mf | $8.5737\lambda_1 +$ | $19.2574\lambda_2 +$ | $80.9079\lambda_3 +$ | $2.5730\lambda_4 +$ | $.7312\lambda_5 +$ | $2.5101\lambda_6 =$ | $-2.8901$ |
| Pd | $26.3314\lambda_1 +$ | $37.0621\lambda_2 +$ | $2.5730\lambda_3 +$ | $104.4952\lambda_4 -$ | $1.5153\lambda_5 +$ | $11.4942\lambda_6 =$ | $7.1099$ |
| HPR | $-\ .8156\lambda_1 -$ | $.6524\lambda_2 +$ | $.7312\lambda_3 -$ | $1.5153\lambda_4 +$ | $.4401\lambda_5 -$ | $1.6429\lambda_6 =$ | $-\ .4515$ |
| Days | $6.9509\lambda_1 +$ | $4.7766\lambda_2 +$ | $2.5101\lambda_3 +$ | $11.4942\lambda_4 -$ | $1.6429\lambda_5 +$ | $76.7578\lambda_6 =$ | $5.8242$ |

ysis," *Journal of Experimental Education*, XXI (1953), pp. 233-41.

3. Hathaway, S.R., and Monachesi, E.D. *Analyzing and Predicting Juvenile Delinquency with the MMPI* (Minneapolis: University of Minnesota Press, 1953).
4. Johnson, P.O. *Statistical Methods in Research* (New York: Prentice-Hall, Inc., 1949).
5. Lubin, A. "Linear and Nonlinear Discriminating Functions," *British Journal of Psychology, Statistical Section*, III (1950), pp. 90-103.
6. Mahalanobis, P.C., and others. "Anthropometric Survey of the United Provinces, 1941: A Statistical Study," *Sankhya*, IX (1939), pp. 89-324.
7. Rao, C. R. "The Problem of Classification and Distance Between Two Populations," *Nature*, CLX (1947), pp. 835-36.
8. Rao, C. R. "Tests of Significance in Multivariate Analysis," *Biometrika*, XXXV (1948), pp. 58-79.
9. Rao, C. R. *Advanced Statistical Methods in Biometric Research* (New York: John Wiley and Sons, 1952).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# ON PROGRAMMING THE INTEGRAL RTH ROOT OF A NUMBER ON FRIDEN STANDARD AND SQUARE ROOT DESK CALCULATORS

WILLIAM J. MOONAN*
U. S. Naval Personnel Research Field Activity
San Diego, California

## 1. Introduction

IT IS THE purpose of this note to indicate some formulas which are useful for finding the roots of real numbers on Friden Calculators. The method used to derive the formulas is due to Newton although its application for taking advantage of the square-root feature on certain Friden Calculators is believed to be new. The mathematics of the derivations are given in the third section of the paper since this information is probably of least interest to those who will operate with the formulas. The second section contains three examples. The cube root example is symbolically programmed for the calculator in section 4. Table I presents a set of formulas which are efficient for finding the rth root of a real number. That is, for finding the square root, cube root, etc., of some number, N. The symbol x in the formulas stands for a guessed or trial value of the root and conceivably may be grossly in error. However, if it is considerably in error, the formulas will ultimately provide an accurate estimate of the root although it will take a longer time than if an estimate only slightly in error were used.

The method to use with the formulas involves essentially selecting a trial value, substituting this (as x) in the appropriate formula and thereby obtaining a new trial value. This new value is again substituted in the formula (as x). This iteration process is repeated until the desired decimal accuracy is obtained.

The formulas in Table I have been arranged so that the arithmetic ultimately involves the accumulation of two quotients. Therefore, the proper machine set-up requires the NON ENT key to be down. Also it is advisable that the ADD key is in the down position. In some of the formulas the divisors of x appear as improper fractions (e.g., $\frac{x}{4/3}$). In actual computations, these numbers can be changed to decimals appropriate to the accuracy required or the problem can be treated as a simple quotient (e.g., 3x/4).

The Square Root Friden enables all roots of the form $r = 2^i$, where $i = 1, 2, 3, \ldots$, of a number to be taken directly without the use of any special formulas. For instance, to obtain the fourth root of 2.0000 we merely find the $\sqrt{2} = 1.41421$ and then find $\sqrt{1.41421} = 1.1892$ which is an accurate estimate of $2^{\frac{1}{4}}$. The higher order roots are difficult to obtain on the standard Friden since the procedure requires the trial values to be raised to large powers. The Square Root Friden is generally more efficient for both the high and low order roots since the quantities involved in the quotients are easier to calculate and moreover the square-root formulas provide estimates which converge faster to the root desired.

## 2. Examples

a) Consider first the problem of taking the cube root of 27. The answer, of course, is 3, but nevertheless let us see how the method works on a problem for which we know the exact answer. Suppose for our example that we wish our answer to be accurate to four decimal places. We set the decimal on the keyboard, lower dial and upper dial to one more than this, 5, and the upper dial decimal is also set to $2 \times 5 = 10$. Recall that the NON ENT and ADD keys are down. Since we will first use a standard Friden (or use the Square Root Friden in the standard manner) and we are seeking a cube root, the proper formula from Table I is:

$$\frac{N}{3x^2} + \frac{x}{1.5} .$$

We first guess a trial value of x = 2.0000, say,

---

*The opinions expressed are solely those of the author and are in no way official; nor are they to be construed as representing those of the U.S. Naval Personnel Research Field Activity or Bureau of Naval Personnel.

### TABLE I

### FORMULAS FOR FINDING ROOTS OF NUMBERS BY ITERATION

| Root | Standard Friden | Square Root |
|---|---|---|
| 2 (Square) | $\frac{N}{2x} + \frac{x}{2}$ | $\sqrt{N} = N^{\frac{1}{2}}$ |
| 3 (Cube) | $\frac{N}{3x^2} + \frac{x}{1.5}$ | $\frac{\sqrt{N}}{1.5\sqrt{x}} + \frac{x}{3}$ |
| 4 (Fourth) | $\frac{N}{4x^3} + \frac{x}{4/3}$ | $N^{\frac{1}{4}}$ |
| 5 (Fifth) | $\frac{N}{5x^4} + \frac{x}{1.25}$ | $\frac{N^{\frac{1}{4}}}{1.25x^{\frac{1}{4}}} + \frac{x}{5}$ |
| 6 (Sixth) | $\frac{N}{6x^5} + \frac{x}{1.2}$ | $\frac{N^{\frac{1}{4}}}{1.5\sqrt{x}} + \frac{x}{3}$ |
| 7 (Seventh) | $\frac{N}{7x^6} + \frac{x}{7/6}$ | $\frac{(xN)^{1/8}}{.875} + \frac{x}{7}$ |
| 8 (Eighth) | $\frac{N}{8x^7} + \frac{x}{8/7}$ | $N^{1/8}$ |
| 9 (Ninth) | $\frac{N}{9x^8} + \frac{x}{1.125}$ | $\frac{N^{1/8}}{1.25x^{1/8}} + \frac{x}{9}$ |
| 10 (Tenth) | $\frac{N}{10x^9} + \frac{x}{10/9}$ | $N^{1/10}$ |

N = Number
x = Trial Value of Root

and copy this value on a worksheet. The number x is entered on the keyboard and also in the multiplier dial. The MULT key is punched and the resulting square is rounded to 5 decimal places and recorded on the keyboard. Then 4.00000 is multiplied by 3 and the product, 12.00000 is entered on the keyboard. The upper dial is cleared by hand. Next the number N = 27 is entered with the twirler knobs in the upper dial around the 10th decimal place and the dividing keys are pressed. Now 2.25000 appears in the lower dial which should now be locked. The upper dial is clear but in general it will not be, so that the CARR CLEAR will have to be punched. We have now found the first of the two quotients. Next x = 2.00000 is entered around the 10th decimal place in the upper dial and 1.50000 (the constant divisor) is entered on the keyboard. After dividing, we find 3.58333 in the lower dial. This number is our second approximation. After copying this on our worksheet we clear the upper dial and lower dial, and unlock the lower dial, and proceed to square 3.58333 and start repeating the evaluation of the formula. After six trials the worksheet would appear as:

| Trial | x |
|---|---|
| 1) | 2.00000 |
| 2) | 3.58333 |
| 3) | 3.08980 |
| 4) | 3.00257 |
| 5) | 2.99999 |
| 6) | 2.99999 |

Because the fifth and sixth trials produce the same number, we cease iterating since a stable trial value has been determined. Therefore, the estimate of the cube root of 27 is 3.0000 which is accurate to four decimal places.

b) The same problem will now be worked using the square-root feature. The appropriate formula from Table I, in this case, is:

$$\frac{\sqrt{N}}{1.5\sqrt{x}} + \frac{x}{3} \quad .$$

We first calculate directly and record the $\sqrt{N} = \sqrt{27} = 5.19615$ and $\sqrt{2} = 1.41421$ which is the square-root of our first trial value. Next 1.41421 is copied on the keyboard and multiplied by 1.5. This product, 2.12132, is copied on the keyboard. After clearing the upper dial by hand, 5.19615 is entered with the twirler knobs in the upper dial around the 19th decimal place. Division gives 2.44948 in the lower dial which should now be locked. Clear the upper dial. Then enter x = 2.00000 around the 10th place in the upper dial. Place 3.00000 (the constant divisor) on the keyboard and divide. The accumulated quotient is 3.11614 and this value is recorded on the worksheet as this is our second trial value. Next we calculate $\sqrt{3.11614}$ and repeat the process. The lower dial should be unlocked when starting a new iteration. The worksheet would appear as:

| Trial | x |
|---|---|
| 1) | 2.00000 |
| 2) | 3.11614 |
| 3) | 3.00108 |
| 4) | 3.00000 |
| 5) | 2.99999 |

$\sqrt{27} = 5.19615$

Notice that one less iteration is required for the same problem by this method than for the method used in sample a).

c) As a final example consider the problem of finding the fifth root of $\pi = 3.14159 = N$, using the Square Root Friden. From Table I the formula to be used is:

$$\frac{N^{\frac{1}{4}}}{1.25x^{\frac{1}{4}}} + \frac{x}{5}$$

The first task is to find $N^{\frac{1}{4}} = (3.14159)^{\frac{1}{4}}$. Since $\sqrt{3.14159} = 1.7724531$, then $N^{\frac{1}{4}} = \sqrt{1.7724531} = 1.33134$. This number is recorded on the worksheet. Because the fourth root of $\pi$ is about 1.3, a good trial value for the fifth root is x = 1.20000. The fourth root of x is now found and multiplied by 1.25. The resulting product, 1.30830, is recorded on the keyboard and 1.33134 is entered around the 10th decimal place in the upper dial. Division gives 1.01761 and the lower dial is locked. Clear the upper dial, enter 1.20000 in the upper dial around the 10th decimal place, and record 5.00000 (the constant divisor) on the keyboard. Division produces 1.25761 in the lower dial and this value is the new x. After four trials, the worksheet appears as:

| Trial | x |
|---|---|
| 1) | 1.20000 |
| 2) | 1.25761 |
| 3) | 1.25726 |
| 4) | 1.25727 |

$(3.14159)^{\frac{1}{4}} = 1.33134$

Therefore, the fifth root of $\pi$ is 1.2573 which is accurate to four decimal places.

## 3. Mathematical Derivations

Newton's method for finding roots of equations is well-known and is described in great detail in many theories of equations and calculus textbooks. Essentially the method says that the roots of an equation in x are approximated by the following expression:

$$x - \frac{f(x)}{f'(x)}$$

where $f(x)$ is the function (or equation) of $x$, the $f'(x)$ is its derivative. The functions of $x$ used in this paper are very simple since we seek only the values of $x$ satisfying the equation $x^r = N$, where $r$ is some positive integer. Of course $f(x) = x^r - N$. We could consider more general functions but the result would be to introduce types of problems not frequently attempted. The derivative of $f(x)$ is $f'(x) = rx^{r-1}$ so that:

$$x - \frac{f(x)}{f'(x)} = x - \frac{(x^r - N)}{rx^{r-1}} = \frac{N}{rx^{r-1}} + \frac{x}{r/(r-1)}$$

Upon substituting $r = 2, 3, \ldots, 10$ in the expression, we produce the formulas under the "Standard Friden" heading of Table I.

It occurred to the author that it may be more efficient to utilize the square-root feature on certain Friden Calculators to find the root of numbers. Accordingly, $f(x)$ was modified to be:

$$f(x) = x^{r/2} - N^{\frac{1}{2}} .$$

Newton's expression leads to:

$$\frac{\sqrt{N}}{\frac{rx}{2}^{(r-2)/2}} + \frac{x}{r/(r-2)}$$

in this case. Substitution of $r = 2$ and $3$ in the above expression gives the formulas for the square and cube roots in the second formula column of Table I. One could also use $r = 4, 5, \ldots$, to generate iterative formulas for other roots, but it appears more wise to use $f(x) = x^{r/4} - N^{\frac{1}{4}}$ for $r = 4, 5$, and $6$. This was done and the expression is:

$$\frac{N^{\frac{1}{4}}}{\frac{rx}{4}^{(r-4)/4}} + \frac{x}{r/(4-4)} .$$

For $r = 7, 8$ and $9$, one might consider $f(x) = x^{1/8} - N^{1/8}$.

The general principle being used is now clear. The Square Root Friden formulas are derived from Newton's expression utilizing a function of the form $f(x) = x^{r/k} - N^{1/k}$ where $k = 2, 4, 8, 16, \ldots$, Newton's formula for the general case is:

$$\frac{N^{1/k}}{\frac{rx}{k}^{(r-k)/k}} + \frac{x}{r/(r-k)} .$$

To find the rth root of N, locate r between two successive k's where k is an integral power of 2. In case r is equally distant between two k's use the smaller k. Substitute k and r in Newton's general formula above and solve for the specific iteration formula. Particular formulas are given in the third column of Table I.

## 4. Programming a Desk Calculator

The art of computation by means of desk calculators has been taught primarily by the demonstration of technique by the more experienced to the less. This unfortunate situation exists because it is usually much easier to show a student how to use a desk calculator than to tell him how, and adequate instructional materials are scarce. Calculator manufacturers have not been too successful in producing manuals which have been useful to novices. Frequently manuals and books on desk calculators have attempted to illustrate computations for particular problems. Consequently, it is often the reader's unpleasant duty to attempt to adapt the principles or steps of the computation (if they can be discovered) to the special task at hand.

Some of the difficulty could be overcome if an efficient code were devised for the instructions. While the author was engaged in writing statistical programs for a digital computer, it occurred to him that instructions similar to those used by computers could be devised for desk calculators. An extensive list of instructions for Friden calculators has been defined. Rather than list these instructions here, suffice it to say that they refer to a) the set-up of the registers and keys of the calculator, b) arithmetic operations, c) the transfer of information, d) branching and decision making, and e) tallying and control. These desk calculator instructions could be given a mnemonic code (such as CAD for "Clear and Add") in the same manner as digital computer instructions, but this does not seem particularly necessary and would impose an additional learning task. The advantage of desk calculator instructions include the increased generality of computation illustration, the compact description of computations, and a permanent and convenient record of procedures. The author believes that the actual demonstration of technique remains the most efficient method of learning desk calculator computation but he also believes the programming of often used and important procedures bears serious consideration, particularly for instruction manuals, and statistical and computation laboratories.

In order to illustrate how the programming might appear, a program for finding the cube root of N is provided in Table II. The reader is invited to follow simultaneously the verbal description in the first example in section 2 of this paper and the commands in Table II. Each command is given a number called a Step Number and a Reference. The Step Number provides the order in which the command is executed. The Reference refers to such things as a set of registers, a constant or variable, pair of numbers in registers, a

### TABLE II

### PROGRAM FOR COMPUTING THE CUBE ROOT OF N BY ITERATION ON A STANDARD FRIDEN CALCULATOR

| Step No. | Command | Reference | Remarks |
|---|---|---|---|
| 00 | STANDARDIZE | | Set calculator in standard condition. |
| 01 | DECIMAL | K;M;A;Q : d+1 | Set registers to d+1 decimals. |
| 02 | DECIMAL | A : 2(d+1) | Set A register to 2(d+1) decimals. |
| 03 | NON ENTRY | DOWN : | Set Non Entry Key down. |
| 04 | ADD | DOWN : | Set Add Key down. |
| 05 | RECORD | x : d+1 | Record approximate cube root, x, on data sheet. |
| 06 | KEYBOARD | x : d+1 | Set Keyboard to x with d+1 decimals. |
| 07 | MULTIPLY | x : d+1 | Multiply by x with d+1 decimals. |
| 08 | ROUND | A : d+1 | Round product in A to d+1 decimals. |
| 09 | TRANSFER | A;K : d+1 | Transfer A register to Keyboard. |
| 10 | MULTIPLY | 3 : d+1 | Multiply by 3. |
| 11 | ROUND | A : d+1 | Round product in A to d+1 decimals. |
| 12 | TRANSFER | A;K : d+1 | Transfer A register to Keyboard. |
| 13 | CLEAR | A, Q : | Clear A and Q registers by hand |
| 14 | SET | A;N : 2(d+1) | Set A register to N around 2(d+1) decimals. Use dial twirlers. |
| 15 | DIVIDE | | Find quotient of A and K. |
| 16 | LOCK | Q : | Lock Quotient register. |
| 17 | CARR. CLEAR | | Clear Carriage. |
| 18 | SET | A; x : 2(d+1) | Set A to x with 2(d+1) decimals. |
| 19 | KEYBOARD | 1.5 : d+1 | Set Keyboard to 1.5. |
| 20 | DIVIDE | | Find Quotient of A and K. |
| 21 | COMPARE | Q; x : d | Compare Q register with last x to d decimals. |
| 22 | BRANCH, UNEQUAL | 26 : | Go to instruction 26 if numbers compared are unequal. |
| 23 | ROUND | Q : d | If equal, round Q to d decimals. |
| 24 | RECORD | x : d | Record $x \doteq N^{1/3}$ on data sheet. |
| 25 | HALT | | Stop calculations. |
| 26 | DEFINE | Q; x : d+1 | Define numer in Q as the new trial x. |
| 27 | BRANCH | 05 : | Go to instruction 05. |

direction on how to set a register or key, or a step number. The Reference is divided essentially into two parts. The first part contains such things as those items just mentioned, and the second part (separated from the first by a colon) always specifies the decimal accuracy or decimal position if it is involved in the command.

The commands have not been specifically defined. By reading the "Remarks" column, the reader should have little difficulty in determining the proper interpretation. Before the reader considers Table II, he should know that the symbols used are defined as: K = Keyboard, M = Multiplier Dial, A = the Upper Dial or Register, Q = the Quotient or Lower Dial, d = the decimal accuracy desired, and x = a trial value of the cube root of N. The command "STANDARDIZE" means that all the dials and registers of the calculator are unlocked and cleared, and all the operational keys are put in their usual (UP) position.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# MUST ALL TESTS BE MULTI-FACTOR BATTERIES?

EDWIN F. ROSINSKI
Medical College of Virginia
Richmond, Virginia

MUCH OF the present day literature on measurement concerns itself with the application and use of multi-factor test batteries. This trend has become so pronounced that journals are publishing series of articles on the use of these batteries.

In one of these series of articles, Super stated that if the detailed data on validity and norms in the Psychological Corporation's Differential Aptitude Tests and the United States Employment Service's General Aptitude Test Battery are as adequate as claimed, then:

> ....the use of singly developed and validated aptitude tests in counseling will in due course be virtually a thing of the past: the multipotentiality of the multifactor test battery gives too many advantages.[1]*

It appears as if Super's statement is not so much a prophecy but a recognition of a reality that is already present. With a pronouncement such as this and others[2] the ultimate question arises: Need single score tests be replaced by multi-factor batteries? It appears, however, that before this question can be answered, there is a need first to investigate the relationship between these single score tests and multi-factor batteries to ascertain if each is performing a different function or the same function. The Psychological Corporation's Differential Aptitude Tests (hereinafter referred to as the D. A. T.) and the World Book Company's Otis Quick Scoring Mental Ability Test: Gamma Form (hereinafter referred to as the Otis) were used to study this relationship. The decision to use these two tests appeared practical for two reasons: (1) the D.A.T. as a multi-factor test is gaining in prominence and use and includes in the manual a great deal of standardization data, and (2) the Otis is one of the most commonly used tests providing a single score IQ.

## Procedure and Results

A sample of 153 students of a senior high-school class was selected. From the permanent record cards of these senior students the appropriate D. A. T. scores and the Otis IQ scores were taken. The former test had been administered to the students when they were in the tenth grade and the latter when the students were in the twelfth grade. There is no apparent reason to suspect that this recognizably selected sample should in any way contaminate the results. Using these raw data from the Otis and the D. A. T. scores, means, standard deviations and Pearson product-moment correlations were obtained.

After these correlations were found, it was necessary to determine partial r's and partial standard deviations.[3] Next the formula for Multiple R was used and a multiple correlation of .82 was computed. Not only does this prove to be significant at the .01 level of confidence but it is an extremely high correlation.

The next step involved the determination of the multiple regression equation or predicting the first variable (Otis) on the basis of the two known variables (D. A. T. Verbal and Abstract Reasoning Raw Scores). With the calculation of the partial regression coefficients, values were substituted in Cornell's[4] multiple regression or prediction equation and the Otis IQ score might then be predicted. This equation reads:

$$X_1 = b_{12.3} X_2 + b_{13.2} X_3 + a$$

where $X_1$ is the predicted Otis score; $b_{12.3}$ and $b_{13.2}$ the partial regression coefficients .51 and .89 respectively; $X_2$ the D. A. T. Verbal Reasoning raw score; $X_3$ the D. A. T. Abstract Reasoning raw score; and "a" the constant value in the equation determined to be 70.5.

Therefore, with the above equation and known values, a hypothetical example can be tested. Assuming that student A has a D. A. T. Verbal and Abstract Reasoning scores of 29 and 34 respectively, what will his predicted Otis IQ be?

Substituting in the formula we have:

$$X_1 = b_{12.3}X_2 + b_{13.2}X_3 + a$$
$$X_1 = .51(29) + .89(34) + 70.5$$

---

*All footnotes will be found at end of article.

TABLE I

A PARTIAL TABLE FOR PREDICTING OTIS (GAMMA) IQ SCORES FROM D. A. T. VERBAL AND ABSTRACT REASONING RAW SCORES*

| D. A. T. Abstract Reasoning Scores | D. A. T. Verbal Reasoning Scores | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| 12 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 108 | 109 | 110 | 111 |
| 13 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 108 | 109 | 110 | 111 |
| 14 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 109 | 110 | 111 | 112 |
| 15 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 109 | 110 | 111 | 112 |

*The table is read in the following manner: If a D. A. T. Verbal Reasoning raw score is 34 and the Abstract Reasoning raw score is 14, the point at which the two columns intersect, the Otis IQ is read: 109.

## TABLE II

### ACTUAL AND PREDICTED OTIS (GAMMA) IQ SCORES WITH DIFFERENCES BETWEEN THE TWO

| Student | Actual Otis IQ | Predicted Otis IQ | Difference | Student | Actual Otis IQ | Predicted Otis IQ | Difference |
|---|---|---|---|---|---|---|---|
| 1 | 104 | 104 | 0 | 42 | 111 | 108 | 3 |
| 2 | 92 | 99 | 7 | 43 | 97 | 104 | 7 |
| 3 | 115 | 108 | 7 | 44 | 99 | 99 | 0 |
| 4 | 96 | 99 | 3 | 45 | 116 | 112 | 4 |
| 5 | 106 | 104 | 2 | 46 | 106 | 104 | 2 |
| 6 | 116 | 108 | 8 | 47 | 121 | 123 | 2 |
| 7 | 101 | 102 | 1 | 48 | 87 | 82 | 5 |
| 8 | 106 | 101 | 5 | 49 | 95 | 79 | 16 |
| 9 | 109 | 108 | 1 | 50 | 105 | 96 | 9 |
| 10 | 103 | 90 | 13 | 51 | 112 | 116 | 4 |
| 11 | 123 | 129 | 6 | 52 | 120 | 115 | 5 |
| 12 | 92 | 83 | 9 | 53 | 106 | 105 | 1 |
| 13 | 110 | 91 | 19 | 54 | 107 | 104 | 3 |
| 14 | 107 | 106 | 1 | 55 | 111 | 106 | 5 |
| 15 | 111 | 108 | 3 | 56 | 93 | 97 | 4 |
| 16 | 87 | 86 | 1 | 57 | 131 | 126 | 5 |
| 17 | 106 | 121 | 15 | 58 | 106 | 113 | 7 |
| 18 | 115 | 114 | 1 | 59 | 96 | 105 | 9 |
| 19 | 97 | 100 | 3 | 60 | 107 | 110 | 3 |
| 20 | 108 | 111 | 3 | 61 | 111 | 111 | 0 |
| 21 | 109 | 110 | 1 | 62 | 119 | 112 | 7 |
| 22 | 110 | 114 | 4 | 63 | 89 | 93 | 4 |
| 23 | 110 | 119 | 9 | 64 | 99 | 101 | 2 |
| 24 | 93 | 88 | 5 | 65 | 115 | 115 | 0 |
| 25 | 100 | 97 | 3 | 66 | 121 | 124 | 3 |
| 26 | 99 | 97 | 2 | 67 | 109 | 118 | 9 |
| 27 | 101 | 105 | 4 | 68 | 111 | 97 | 14 |
| 28 | 107 | 111 | 4 | 69 | 88 | 90 | 2 |
| 29 | 115 | 115 | 0 | 70 | 90 | 97 | 7 |
| 30 | 92 | 102 | 10 | 71 | 125 | 112 | 13 |
| 31 | 116 | 120 | 4 | 72 | 104 | 109 | 5 |
| 32 | 109 | 116 | 7 | 73 | 92 | 105 | 13 |
| 33 | 117 | 119 | 2 | 74 | 99 | 106 | 7 |
| 34 | 123 | 121 | 2 | 75 | 99 | 113 | 14 |
| 35 | 100 | 82 | 18 | 76 | 127 | 91 | 8 |
| 36 | 107 | 118 | 11 | 77 | 83 | 91 | 8 |
| 37 | 100 | 113 | 3 | 78 | 99 | 105 | 6 |
| 38 | 98 | 99 | 1 | 79 | 113 | 111 | 2 |
| 39 | 121 | 115 | 6 | 80 | 96 | 97 | 1 |
| 40 | 102 | 101 | 1 | 81 | 112 | 118 | 6 |
| 41 | 110 | 104 | 6 | 82 | 119 | 122 | 3 |

## TABLE III

### NUMBER AND PERCENT OF PREDICTIONS MADE WITHIN CERTAIN DIFFERENCES BETWEEN ACTUAL AND PREDICTED IQ'S

| Diff. Between Actual and Predicted IQ | No. of Predictions | Percent of Predictions | Cumulative Percentage |
|---|---|---|---|
| 0 | 5 | 7 | 7 |
| 1 | 10 | 12 | 19 |
| 2 | 19 | 11 | 30 |
| 3 | 11 | 13 | 43 |
| 4 | 8 | 10 | 53 |
| 5 | 8 | 10 | 63 |
| 6 | 5 | 7 | 70 |
| 7 | 8 | 10 | 80 |
| 8 | 2 | 2 | 82 |
| 9 | 5 | 7 | 89 |
| 10 | 1 | 1 | 90 |
| 11 | 1 | 1 | 91 |
| 12 | 0 | 0 | 91 |
| 13 | 3 | 3 | 94 |
| 14 | 2 | 2 | 96 |
| 15 | 1 | 1 | 97 |
| 16 | 1 | 1 | 98 |
| 17 | 0 | 0 | 98 |
| 18 | 1 | 1 | 99 |
| 19 | 1 | 1 | 100 |

## TABLE IV

### MEANS, STANDARD DEVIATIONS AND DIFFERENCES IN PREDICTED AND ACTUAL IQ SCORES

| | Actual | Predicted | Difference |
|---|---|---|---|
| Mean | 106.0 | 106.2 | 0.2 |
| Standard Deviation | 9.8 | 10.1 | 0.3 |

$$X_1 = 14.79 + 30.26 + 70.5$$
$$X_1 = 115.55$$

Rounded off to the nearest hundred, we have student A's predicted Otis IQ as 116.

Nevertheless, the task of predicting presented another problem. In attempting to predict several Otis IQ's it quickly became obvious that such a procedure of substituting in a formula was suitable with a small number of cases. On the other hand, if this was to be accomplished for a large group, it would be much too time consuming. Consequently, a prediction table was constructed by which the Otis score could be easily determined once the D. A. T. Verbal and Abstract Reasoning scores were known. Such a table was completed, but since it is large (49 by 47 columns) only a portion is here reproduced (Table I).

To test the prediction equation or, as it developed in this case, the prediction table, a new sample of students was selected. An entire senior class of 82 students in the same school where the original correlations were made was used. Again from the permanent record cards of these students the D. A. T. Verbal and Abstract Reasoning and the Otis scores were transcribed on a master work sheet. From the D. A. T. data the predicted IQ's were made and compared with the actual IQ's.

While the task of developing and recording the figures for the prediction table was tedious, the effort was rewarded by the ease with which the predicted scores were made.

Table II represents the findings of this procedure of predicting scores: the actual IQ scores along with the predicted IQ scores and the differences between the two. Prior to reading the table it is important to remember that one standard error of estimate for this prediction is 4.3 and the usual procedure is to expect, by the definition of a standard error of estimate, approximately 90% of the predictions within two standard errors of estimate, that is 90% of the predictions should fall within 8.6 (9) points of the predicted IQ. Table III shows the number and percent of predictions made within a certain difference.

From these two tables it is obvious that the predictions are not all perfect. This might well be expected; for perfect correlations, while highly desirable, are extremely unlikely. Yet, it does appear significant when considering the number of predictions that fell within the two standard errors of estimate and the reliability of the Otis itself. In this case, the prediction table and/or prediction equation appear to have some value.

Perhaps citing another figure might be more illustrative of the value of this prediction, for studies of such a nature as this are primarily concerned with mean scores and mean values. Therefore, when comparing the mean IQ's and standard deviations of the predicted and actual scores in Table IV, the accuracy of the predictions seems more vivid.

The greatest value, perhaps, of this study lies in this table. It appears almost extraordinary that predicting group mean IQ's could be accomplished with such a degree of accuracy as illustrated here.

## Summary

In correlating the D. A. T. Verbal and Abstract Reasoning subtests with the Otis, a multiple correlation coefficient of .82 was found. From the multiple correlation coefficient, the prediction equation was developed and predicted 90% of the cases within two standard errors of estimate. While all the predictions were not perfect, the Multiple R was almost identical to the Otis reliability coefficient. In addition, while differences in individual predictions were noted, the predicted mean group IQ's were as close to the actual IQ's as possible (106 and 106.2).

## Conclusions

From the data presented in this study, several conclusions seem appropriate.

1. Because of the significant and high Multiple R (.82) it appears as if the D.A.T. Verbal and Abstract Reasoning subtest are performing the same function as the Otis, i.e., measuring intelligence.
2. Because of the significant and high Multiple R, it appears as if the D.A.T. Verbal and Abstract Reasoning subtests can predict Otis IQ scores.
3. In actually predicting Otis IQ's on the basis of D.A.T. Verbal and Abstract Reasoning subtests, approximately 90% of the predictions will fall within nine points.
4. Predicting mean Otis IQ's for groups can be accomplished with considerable accuracy and confidence.

## Implications

Several implications manifest themselves as a result of this study. If the multi-factor tests measure the same factors as some single score tests, then it only seems logical that they might ultimately replace these single score tests, but by the same token if they are measuring the same factors as single score tests, why not continue using these single score tests? It appears that even though the multi-factor tests can, with some degree of accuracy, do the same task as single tests the rush toward these multi-factor tests should be a considered one until the predictive value of these tests is substantiated with proven data. This is especially pertinent when considering the amount of time that each test takes. The Otis requires thirty minutes testing time while the D. A. T. requires approximately three hours, not to mention the time required for directions and scoring.

An argument for multi-factor test batteries could be the possible multipotentiality of the multi-factor tests, i. e., that the various subtests could be recombined to predict other phenomena.

What it seems to boil down to is this: that both tests tend to measure the same factors, but because they measure the same factors it does not necessarily follow that one should replace the other. Test users perhaps should consider some gross criteria before deciding on a particular test: (1) the purpose behind using a particular test, (2) the time allotted to accomplish the particular purpose, and (3) what the testor and testee are most comfortable in handling.

FOOTNOTES

1. Donald E. Super, "The Use of Multifactor Test Batteries in Guidance," Personnel and Guidance Journal, XXXV (September 1956), p. 14.
2. 1955 Invitational Conference on Testing Problems, sponsored by the Educational Testing Service of Princeton, New Jersey.
   Robert D. North, The Use of Multi-Factor Aptitude Tests in School Counseling Proceedings of the 1955 Invitation Conference on Testing Problems, New York, October 29, 1955, pp. 11-15.
3. Any standard test on statistics in psychology and education provides formulae for these values.
4. Francis G. Cornell, The Essentials of Educational Statistics (New York: John Wiley and Sons, 1956), p. 320.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# AN EVALUATION OF THE LABORATORY IN A COLLEGE GENERAL BIOLOGY COURSE

DOUGLAS M. DEARDEN
University of Minnesota

THE RESEARCH reported to date in the literature indicates that no one type of laboratory is clearly superior in developing all of the outcomes desired from the science laboratory. The acceptable science studies indicate, for example, that the conventional type of laboratory has proven superior to other laboratory methods in teaching laboratory skills, but not in promoting the acquisition of facts.

The research reported in this study was conducted to help determine the contribution of several types of biology laboratories in a college biology course designed for freshmen and sophomores who do not plan to major in a science.

## Experimental Treatments

The four experimental treatments compared were: (1) individual (conventional) laboratory, (2) demonstration laboratory, (3) workbook exercises, and (4) a biological term paper.

The individual laboratory work was based on the laboratory work found in many of the published college laboratory manuals, and involved a weekly two-hour laboratory period. The demonstration laboratory required one laboratory hour per week. The laboratory instructors had set up experiments and materials prior to the laboratory period for the students to observe and study. This demonstration type of laboratory eliminated many of the more routine aspects of the laboratory work for the student and allowed them to concentrate on the important points and principles.

The workbook exercises were submitted weekly and were based on questions in the demonstration laboratory manual. The answers to these exercises had to be obtained from non-laboratory sources, such as texts. The biological report was used as an experimental control.

The lecture portion of the class was presented by one lecturer who kept the lectures as constant as possible for all sections of the class.

## Measuring Instruments

Three tests were used to evaluate the achievement of students randomly assigned to the four experimental treatments. The students took these tests before beginning the course, upon completion of the course, and three months after completing the course. The three tests were tests of Biological Knowledge, Scientific Thinking, and Biological Attitudes. These tests were validated and the reliability coefficients for the tests were computed. The reliability coefficients were computed by the split-halves method, using the maximum likelihood estimate of the correlation coefficient and then correcting this coefficient with the Spearman-Brown formula. The corrected reliability coefficients were 0.87 for the Biological test, 0.86 for the Biological Attitude test and 0.85 for the Scientific Thinking test.

## Experimental Population

The students involved in the experiment were the 924 students who completed the one quarter General College Human Biology course, G.C. 10-A during the 1958-59 academic year. These students were enrolled in the General College, Home Economics, Education, Mortuary Science, and the College of Science, Literature, and the Arts. The students in the six sections of this course were randomly assigned in approximately equal numbers to the treatments through the use of a table of random numbers.

## Experimental Design

A stratified randomized design was used in this study, with the experimental treatments being randomly assigned over all of the experimental material. The students were stratified for college and sex prior to being assigned to treatments as a means of more equally distributing the students of various abilities.

The desirable statistical tool to test the null hypotheses of the study was the analysis of variance. Of the two main models of this analysis, the fixed effects model, Model I, is designed for studies such as the present study. T-tests, Duncan's Multiple Range Test, and the Johnson-Neyman Technique were also included in the statistical analyses.

## Experimental Results

A comparison of the initial and final test scores by means of t-tests indicated that each of the four treatments had significantly increased student's scores on the three tests. The treatment which most consistently increased the variance on the final tests was the individual laboratory, which suggests that this treatment allowed more adequately for individual differences in student academic ability.

The overall results from the experiment are presented in Tables I, II, and III which give the means of the tests given at the end of the experimental period for each of the six sections.

When the final Biological Knowledge Test scores were analyzed by a two-way analysis of variance (treatment and ability level) none of the treatments was significantly better in increasing student scores. Table IV presents the analysis of variance results for the lecture section which most closely approached significant results between treatments on this test.

The conclusion which resulted from the analysis of the Biological Knowledge test was that none of the four treatments was significantly better in helping students learn the factual aspects of biology.

The Scientific Thinking tests were similarly analyzed by the analysis of variance technique to determine which treatment functioned best in improving scientific thinking by students. The differences in achievement on this test were not statistically significant, indicating that none of the treatments was superior in improving scientific thinking by students.

When the Biological Attitude Test results were analyzed, significant differences were found between the treatment groups in two of the lecture sections. The significant differences in section one, Fall quarter (Table V) were further analyzed by Duncan's Multiple Range Test, which indicated that the individual laboratory group of this section had significantly higher attitude test scores than the other three treatments. In section three, Fall quarter, the workbook group was significantly higher on this test than the other three treatments. However, the total results do not indicate that any one treatment was consistently superior in improving student's scores on this attitude test.

The preceeding analysis of variance results also were used to determine if any treatment was superior so far as being more suitable for students of comparatively high, middle, or low academic ability. The chosen criterion of academic ability was the student's ACE scores. The analysis of variance results indicated that the students of a given ability group achieved equally well on the tests no matter which treatment they were assigned to. Thus none of the treatments proved to be better for the students of any of these three comparative levels of ability.

## Analysis of the Follow-up Tests

Three months after completing the experimental course, a random sample of students again took the tests used in this study. Of the 122 students selected, 79.5 percent took the three tests. T-tests for correlated measures were used to compare the final and follow-up test means and variances.

The students in all treatment groups tended to score lower on the follow-up tests of Biological Knowledge and Attitudes. However, each of the treatment groups increased their scores on the follow-up Scientific Thinking Test. In comparing the variances of the final and follow-up tests, the individual laboratory group most consistently increased the variability in scores on the follow-up tests, indicating that this treatment allowed more adequately for individual differences in learning.

A two-way analysis of variance of the follow-up tests (treatments × number of additional biology courses taken by students) indicated that the four treatments were not significantly different in promoting a more permanent learning by students of biological attitudes, facts, or scientific thinking. This conclusion held both for those students who did or did not take an additional biology course after completing the experimental course.

## Johnson-Neyman Technique*

While no significant differences were found between the treatments on the Biological Knowledge and Scientific Thinking Tests, some differences were noted on the Biological Attitude Test. Thus it was advisable to collate the information from all of the six single experiments to see if the combined data indicated any significant differences between the treatments. A qualitative modification of the Johnson-Neyman Technique was used to combine and collate the data from the six separate experiments.

In this analysis, two treatments are compared at one time. For example, when the combined demonstration laboratory groups were compared with the combined individual laboratory groups on the Attitude test, the difference (D) in mean final Attitude test scores between the two treatments and the variance ($S_D^2$) of D were computed. Then the value of $t_O$ was calculated, where

$$t_O = \frac{D}{S_D^2} .$$

Referring to the table of t, with the appropriate degrees of freedom, the null hypothesis of no mean differences between the two groups could be re-

---

* The technical background for this technique is presented in the text by Palmer O. Johnson and Robert Jackson, Modern Statistical Methods: Descriptive and Inductive (Chicago: Rand McNally and Co., 1959), 427-8.

## TABLE I

### MEAN SCORES OF STUDENTS ON THE FINAL BIOLOGICAL KNOWLEDGE TEST

| Treatment | Lecture Section<br>Sec. 1 Fall | Sec. 2 Fall | Sec. 3 Fall | Sec. 1 Winter | Sec. 2 Winter | Sec. 1 Spring |
|---|---|---|---|---|---|---|
| Demonstration Lab. | 54.76 | 50.04 | 52.92 | 50.64 | 55.50 | 48.81 |
| Workbook | 52.31 | 52.06 | 55.33 | 51.29 | 53.36 | 53.05 |
| Report | 51.44 | 49.75 | 53.67 | 49.69 | 52.60 | 50.42 |
| Individual Lab. | 56.40 | 53.13 | 56.42 | 46.13 | 52.83 | 50.62 |

## TABLE II

### MEAN SCORES OF STUDENTS ON THE FINAL SCIENTIFIC THINKING TEST

| Treatment | Lecture Section<br>Sec. 1 Fall | Sec. 2 Fall | Sec. 3 Fall | Sec. 1 Winter | Sec. 2 Winter | Sec. 1 Spring |
|---|---|---|---|---|---|---|
| Demonstration Lab. | 32.04 | 30.20 | 28.82 | 30.73 | 32.97 | 34.52 |
| Workbook | 32.19 | 31.50 | 31.33 | 32.26 | 33.41 | 31.62 |
| Report | 30.75 | 31.78 | 29.84 | 30.00 | 32.29 | 32.58 |
| Individual Lab. | 35.53 | 30.00 | 29.75 | 28.33 | 29.50 | 35.90 |

## TABLE III

### MEAN SCORES OF STUDENTS ON THE FINAL BIOLOGICAL ATTITUDE TEST

| Treatment | Lecture Section<br>Sec. 1 Fall | Sec. 2 Fall | Sec. 3 Fall | Sec. 1 Winter | Sec. 2 Winter | Sec. 1 Spring |
|---|---|---|---|---|---|---|
| Demonstration Lab. | 11.27 | 11.58 | 11.82 | 13.73 | 12.81 | 12.57 |
| Workbook | 12.35 | 12.56 | 14.15 | 13.76 | 13.23 | 13.19 |
| Report | 12.00 | 9.90 | 13.86 | 12.38 | 14.07 | 15.46 |
| Individual Lab. | 15.87 | 14.47 | 12.75 | 12.60 | 13.22 | 13.19 |

TABLE IV

ANALYSIS OF VARIANCE OF THE BIOLOGICAL KNOWLEDGE TEST
(Section One Fall Quarter)

| Source of Variation | D. F. | Sum of Squares | Mean Square | F-Test |
|---|---|---|---|---|
| Between Ability Levels (i) | 2 | 4169.8974 | 2084.9487 | $F = 16.32$ $P < .05$ Reject the Ho of equal means among ability levels. |
| Between Treatments (j) | 3 | 448.8613 | 149.6204 | $F = 1.17$ $P > .05$ Do not reject the Ho of equal treatment means. |
| Interaction (i) × (j) | 6 | 1651.5720 | 275.2620 | $F = 2.15$ $P > .05$ Do not reject the Ho of no interaction. |
| Error | 144 | 18390.5668 | 127.7123 | |
| Total | 155 | 24660.8975 | | |

## TABLE V

### ANALYSIS OF VARIANCE OF THE ATTITUDE TEST
(Section One Fall Quarter)

| Source of Variation | D. F. | Sum of Squares | Mean Square | F-Test |
|---|---|---|---|---|
| Between Ability Levels (i) | 2 | 46.0385 | 23.0193 | F = 0.88 $P > .05$<br>Do not reject the $H_0$ of equal means among the ability levels. |
| Between Treatments (j) | 3 | 243.1798 | 81.0599 | F = 3.10 $P < .05$<br>Reject $H_0$ of equal treatment means. |
| Interaction (i × j) | 6 | 82.0198 | 13.6700 | F = 0.52 $P > .05$<br>Do not reject the $H_0$ of no interaction. |
| Error | 144 | 3763.4543 | 26.1351 | |
| Total | 155 | 4134.6924 | | |

## TABLE VI

### COMPARISON OF EXPERIMENTAL TREATMENTS USING THE JOHNSON-NEYMAN TECHNIQUE

| Test* | Treatments Compared | D. F. | D | $S_D$ | $t_o$ | Probability of $t_o$ |
|---|---|---|---|---|---|---|
| A | Report - Individual Lab. | 374 | -0.72 | 0.70 | 1.02 | $.3 < P < .4$ |
| | Workbook - Individual Lab. | 371 | -0.48 | 0.64 | 0.75 | $.4 < P < .5$ |
| | Workbook - Report | 536 | 0.53 | 0.50 | 1.06 | $.2 < P < .3$ |
| | Demonstration Lab. - Individual Lab. | 372 | -1.09 | 0.64 | 1.70 | $.05 < P < .1$ |
| | Demonstration Lab. - Report | 537 | -0.20 | 0.50 | 0.41 | $.6 < P < .7$ |
| | Demonstration Lab. - Workbook | 534 | -0.73 | 0.47 | 1.55 | $.1 < P < .2$ |
| B | Report - Individual Lab. | 374 | -0.03 | 0.97 | 0.03 | $.9 < P < 1.0$ |
| | Workbook - Individual Lab. | 370 | 0.31 | 0.96 | 0.33 | $.7 < P < .8$ |
| | Workbook - Report | 535 | 0.03 | 0.72 | 0.04 | $.9 < P < 1.0$ |
| | Demonstration Lab. - Individual Lab. | 372 | 0.42 | 1.00 | 0.41 | $.6 < P < .7$ |
| | Demonstration Lab. - Report | 537 | -0.36 | 0.74 | 0.49 | $.6 < P < .7$ |
| | Demonstration Lab. - Workbbok | 533 | -0.54 | 0.73 | 0.73 | $.4 < P < .5$ |
| S | Report - Individual Lab. | 374 | 0.69 | 1.13 | 0.60 | $.5 < P < .6$ |
| | Workbook - Individual Lab. | 371 | 0.66 | 1.10 | 0.60 | $.5 < P < .6$ |
| | Workbook - Report | 536 | 0.37 | 0.83 | 0.44 | $.6 < P < .7$ |
| | Demonstration Lab. - Individual Lab. | 372 | 0.47 | 1.08 | 0.44 | $.6 < P < .7$ |
| | Demonstration Lab. - Report | 537 | -0.07 | 0.82 | 0.09 | $.9 < P < 1.0$ |
| | Demonstration Lab. - Workbook | 534 | -0.47 | 0.80 | 0.59 | $.5 < P < .6$ |

* A = Attitude final test
B = Biological knowledge final test
S = Scientific thinking final test

jected or accepted.

Using this technique to compare each of the combined treatments with every other combined treatment, the conclusion was drawn that none of the mean test scores were significantly higher at the .05 level of significance. Thus, at this level, none of the treatments was significantly better in improving the mean final scores of students on these three tests. However, at the .10 level, the individual laboratory, when compared with the demonstration laboratory, was associated with a significant increase in scores on the Biological Attitude Test. A summary of the results of all of these comparisons is presented in Table VI.

A sample of the experimental groups of students was compared with two random samples of students who had completed this course several years ago. The ACE scores of these samples of students were compared by an analysis of variance technique. The experimental group was not significantly different, with regard to ACE scores, from the students of previous years. Thus the results of this experiment can be generalized to that population of students taking the experimental course, assuming that the characteristic of future students do not change.

## Summary

Four experimental treatments (individual or conventional laboratory, demonstration laboratory, workbook exercises, and a biological term paper) were statistically evaluated in a college general biology course. Tests of Biological Knowledge, Scientific Thinking, and Biological Attitudes were used to compare the achievement of students randomly assigned to these treatments.

A significant increase in test scores was associated with each treatment at the end of the instructional period, but none of the treatments was superior to the other treatments in increasing student's scores. Each of the treatments allowed for individual differences in the academic ability of students, but the individual laboratory allowed more adequately for these differences in learning ability and thus allowed the students to learn at a rate more consistent with their ability.

In comparing the test results of students with comparatively high, middle, and low academic ability, none of the treatments was found to be more suited to students of a particular level of ability.

At the .10 level, the Johnson-Neyman Technique indicated a difference between two of the treatments. When the individual laboratory and demonstration laboratory groups were compared, the individual laboratory groups had significantly higher mean test scores on the Biological Attitude Test.

None of the treatments was significantly superior in promoting a more permanent learning of the course materials, as measured by tests administered three months after the students had completed the experimental course.

The results of experimentation can provide useful practical results. Based on the results of this experiment, a combination of these methods could be used to good advantage. The advantages of all four treatments could be incorporated into a course, with particular emphasis being placed on the individual laboratory because of its better adjustment to individual differences in learning ability of students. In addition, if the four treatments were carefully integrated into a course, student interest in the course materials could be increased.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# APPLICATION OF MATRICES IN THE ANALYSIS OF SOCIOMETRIC DATA

JOHN X. JAMRICH
Michigan State University
East Lansing, Michigan

A NUMBER OF research workers in the field of sociometry have clearly identified the inadequacy of the ordinary methods used in the analysis of sociometric data. Briefly, the inadequacies stem from the fact that ordinary methods of analysis do not determine, uniquely, status levels of the individuals or of sub-groups of individuals in the group being studied. Furthermore, present methods lack specific quantitative characteristics to permit analysis of change in the social structure of a group, in the relative status of an individual within the group, or of the sub-groups, usually referred to as cliques.

Katz* has presented a mathematical derivation of a "status vector" of a given group. The derivation of the numerical elements of this vector is a relatively cumbersome process, involving the solution of "n" simultaneous equations, where "n" is the number of individuals comprising the group.

It is the purpose of this paper to present the development of a relatively simple numerical characteristic for sociometric data which will permit:

1. determination of the status of an individual within the group, taking into account his choices and the kinds of choices by others in which he is involved;
2. determination of a characteristic vector for the entire social group so that it may be compared with other groups and with sociometric data for the same group at a later date;
3. the identification of cliques, and their relative status within the group, and changes which may be taking place from time to time as new sociometric data are gathered for the same group.

The understanding of the development and computation is based on the simple concept of addition and multiplication of matrices. By definition, the sum of two matrices:

$$(1)\quad M_1 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ and } M_2 = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

is given by

$$(2)\quad M_1+M_2 = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

The product of two matrices, with elements as above, is defined as:

$$(3)\quad M_1 \circ M_2 = \begin{pmatrix} a_{11} b_{11} + a_{12} b_{21} & a_{11} b_{12} + a_{12} b_{22} \\ a_{21} b_{11} + a_{22} b_{21} & a_{21} b_{12} + a_{22} b_{22} \end{pmatrix}$$

Thus, if we have a matrix

$$(3a)\qquad M_1 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

then, $M_1^2$ would be $M_1$ multiplied by itself, namely:

$$(4)\qquad M_1^2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \circ \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

and

$$(5)\quad M_1+M_1^2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}$$

For later references, it should be pointed out

---

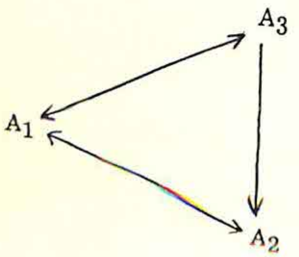
*Leo Katz. "A New Status Index Derived from Sociometric Analysis," *Psychometrika*, XVIII (March 1953), pp. 39-43.

that the sums of the columns of the matrix $M_1 + M_1^2$ are

$$(5, 5, 3). \tag{6}$$

Consider, as an example, the following simple sociogram involving three individuals, $A_1$, $A_2$, and $A_3$, each individual having made one or two choices of equal "strength."

(7)



The matrix representing this sociogram will be:

$$M_1 = \begin{array}{c|ccc} & A_1 & A_2 & A_3 \\ \hline A_1 & 0 & 1 & 1 \\ A_2 & 1 & 0 & 0 \\ A_3 & 1 & 1 & 0 \end{array} \tag{8}$$

In the above matrix (8), the entries in the columns indicate that $A_1$ is chosen by certain of the other individuals, and the entries in the rows indicate that $A_1$ has chosen certain of the other people in the group. In other words, the entries represent the number of one-stage relations (or communications) between the individuals.

The above matrix (8), is seen to be precisely the one used as an illustration in (3a). Now, let us multiply the matrix in (8) by itself. We obtain:

$$M_1^2 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \tag{9}$$

If we consider $M_1$ to be a communication matrix, then in the square of it, $M_1^2$, the entry in the ith row, jth column gives the number of "2-stage" communications between $A_i$ and $A_j$. For example, $A_3$ can communicate once with $A_1$ in two stages, while $A_2$ cannot communicate with $A_1$ in two stages. We might say, then, that the entries represent the number of two-stage relations among the corresponding individuals.

Now forming the sum, $M_1 + M_1^2$, we obtain the matrix (10) shown at the top of column two on this page.

$$M_1 + M_1^2 = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix} \tag{10}$$

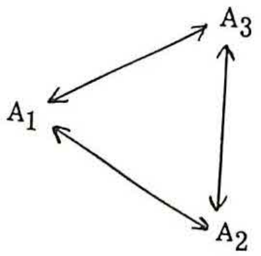
The column sums of (10) are (5, 5, 3)

From the interpretation of (8) and (9), we see that the entries in (10) represent the number of one- and two-stage relationships between corresponding individuals. Hence, the column sums will give the total number of one- and two-stage "inward" relations (including the one with himself) enjoyed by each individual. This latter fact is used here as an indicator of the status of the individual within the group. (The row sums can be interpreted as "outward" one- and two-stage relations or communications.)

Now, a consideration of the matrix in (10) and the original sociogram reveals some interesting facts. First, the diagonal of the matrix in (10) gives, for each of the individuals, the number of "mutual choices" enjoyed by each one. Thus, $A_1$ has two mutual choices, while $A_2$ and $A_3$ have only one each. As indicated above, the column sum vector (5, 5, 3), gives the relative status of each individual within the group. Thus, according to this characteristic vector, $A_1$ and $A_2$ have equal status, while $A_3$ has less than either of the others.

Using this same group of three individuals, let us illustrate the manner in which the column-sum vector can be used to compare change in a given group from one time to another. For this purpose, let us suppose that a subsequent application of the same procedure yielded the following sociogram:

(11)



whose matrix is,

$$M_1 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

The square of this matrix is

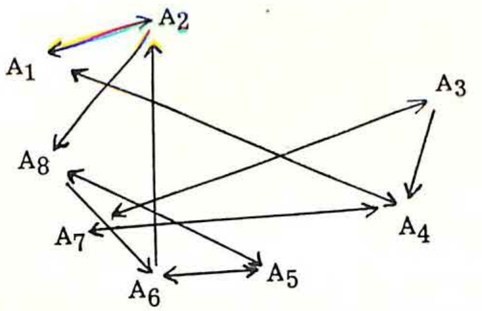
$$M_1^2 = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \tag{12}$$

and the sum,

$$M_1 + M_1^2 = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

The column-sum vector is (6, 6, 6). In other words, compared with the column-sum vector of the sociogram of (7), it is evident that all three of the individuals enjoy two mutual choices, as revealed by the diagonal values in (12), and, they are all of equal status.

Remembering that these illustrations assume two choices for each individual, and two choices of equal strength, let us illustrate the utilization of the method with a more complex group, say, of eight individuals. First, let us consider the following sociogram for these eight:



The matrix for this sociogram would be

$$(13) \qquad C = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

and the square of it is

$$(14) \qquad C^2 = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 1 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

The sum, $C + C^2$ is shown at the top of the next column, and the resulting column-sum vector is

$$(16) \qquad V\!: \quad \begin{matrix} (7, 6, 3, 8, 6, 6, 6, 6) \\ (A_1\ A_2\ A_3\ A_4\ A_5\ A_6\ A_7\ A_8) \end{matrix}$$

From the values of the diagonal in (14), it is clear that $A_1$, $A_4$, $A_5$, and $A_7$ each have two mutual choices to their credit. From the characteristic column-sum vector, it is seen that $A_4$ has the "most" status, which fact can be substantiated from the original sociogram, $S_1$, for the purposes of this discussion. Similarly, $A_1$ is seen to have the next highest status designation, followed by $A_5$, $A_6$, $A_7$ and $A_8$. The latter four are indicated by the vector sums as being of equal status, a fact which can again be checked in the original sociogram.

$$(15) \qquad C + C^2 = \begin{pmatrix} 2 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 2 & 0 & 0 & 2 & 0 \\ 1 & 1 & 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 & 2 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 2 \\ 1 & 0 & 1 & 2 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 2 & 2 & 0 & 1 \end{pmatrix}$$

It may be worth noting that the column-sum vector values of $A_1$ and $A_5$ are 7 and 6 respectively. From the sociogram, one would conclude that these two individuals have equal status on the basis of the fact that each enjoys two mutual choices. The values of 7 and 6 in the column-sum vector indicate the fact that $A_1$ has one more two-stage "inward" relation in the group than does $A_5$, a fact which may be of some importance in the inter-group relationships.

The interpretation of the column sums might be illustrated at this point by showing the meaning of the number "7" which is the value associated with $A_1$. The seven "inward" one- and two-stage relations for $A_1$ are as follows:

$$A_1 \longleftarrow A_2 \longleftarrow A_6$$
$$A_1 \longleftarrow A_2 \longleftarrow A_1$$
$$A_1 \longleftarrow A_4 \longleftarrow A_1$$
$$A_1 \longleftarrow A_4 \longleftarrow A_3$$
$$A_1 \longleftarrow A_4 \longleftarrow A_7$$
$$A_1 \longleftarrow A_2$$
$$A_1 \longleftarrow A_4$$

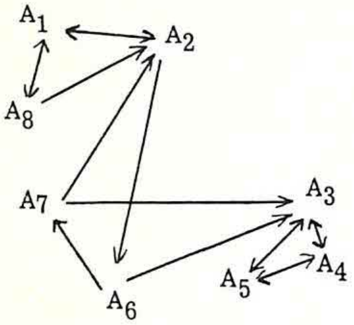
By using the characteristic column-sum vector to identify the individuals with the most status, and using the original matrix, we are able to identify the cliques in this group and associate with them a status designation by using the column-sum vector for each individual in the clique.

For example, from (16) we see that $A_4$ has the highest column sum. From the diagonal entry in (14) we see that $A_4$ enjoys two mutual choices which, from (13) are with $A_1$ and $A_7$. Hence, $A_4$, $A_1$, and $A_7$ can be said to form a clique.

Similar analyses yield the following cliques, the first three of which actually constitute a large sub-group:

$$(17) \qquad \begin{array}{lllcl} A_4 & A_1 & A_7 & : & (8,\ 7,\ 6) \\ A_4 & A_3 & A_7 & : & (8,\ 3,\ 6) \\ A_1 & A_2 & A_4 & : & (7,\ 6,\ 8) \\ A_5 & A_6 & A_8 & : & (6,\ 6,\ 6) \end{array}$$

Now consider the same eight individuals with the following sociogram at some later time.



The matrix representing this sociogram is

$$(18) \qquad D = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The square of this matrix is

$$(19) \qquad D^2 = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

and

$$(20) \qquad D + D^2 = \begin{pmatrix} 2 & 2 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

The characteristic column-vector for (20) is

$$(21) \qquad \begin{array}{cccccccc} (6, & 7, & 10, & 8, & 8, & 4, & 2, & 3\ ) \\ (A_1, & A_2, & A_3, & A_4, & A_5, & A_6, & A_7, & A_8) \end{array}$$

and the cliques, with their characteristic column-sum vectors, are

$$\begin{array}{lcl} A_1\ A_2\ A_8 & : & (6,\ 7,\ 3) \\ A_3\ A_4\ A_5 & : & (10,\ 8,\ 8) \end{array}$$

It is evident that the individual with the most status is $A_3$, quite a change from the case in (13). The status of the other individuals can be read quite easily from (21). There are two strong cliques, the stronger of them being $A_3$, $A_4$, $A_5$.

## Summary

The use of the characteristic column-vector presented above appears to provide a useful method for quantifying sociometric data. Specifically, it affords a relatively simple procedure for determining a numerical value for the status of individuals in a group.

It should be emphasized, however, that the method presented above utilizes the notion of the incoming, one- and two-stage relations enjoyed by each individual in the group. This procedure, then, places additional emphasis not only upon how many chose a given individual, but whether those who chose him were chosen by others in the group.

Finally, the utility of the method of the paper will depend upon additional research with sociometric data to ascertain the significance of the two-stage relationships in interpersonal choices of individuals.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# A TABLE TO FACILITATE THE CALCULATION OF NORMALIZED T-SCORES FOR SMALL SAMPLES (N>5<51)

PETER F. MERENDA
Walter V. Clarke Associates, Inc.
East Providence, Rhode Island

SEVERAL AUTHORS of psychological measurement and statistics texts, Garrett (1) and Edwards (2), for example, have included in the appendix a table to facilitate the calculation of T-scores. The use of this table, however, requires the performing of all of the steps in the usual procedure for normalizing distributions and calculating T-scores.

The table presented in this article requires only that the first four steps in the usual procedure be performed. The final step of calculating percentages, which is the most time-consuming of all the steps, before entering the T-score table is eliminated. In order to enter this table, all one needs to know are the number in the sample and cumulative frequency below a given score plus one half of the frequency of the given score. The table is divided into two parts: Table I for N's between 6 and 24 inclusive, and Table II for N's of 25-50 inclusive. The procedure and instructions for use with this table are given below with example:

## INSTRUCTIONS FOR USE WITH T-SCORE TABLE

| Step | Procedure |
|---|---|
| 1. | Prepare frequency distribution of scores. |
| 2. | Cumulate scores, i. e., add each successive row of scores. |
| 3. | Tabulate the frequency below a score + 1/2 the frequency of a given score. |
| 4. | Enter T-score table with N (number in sample) and read standard normalized scores in the appropriate cell for values calculated in step 3. |

Example:

N = 10

| X | f | Cum. f. | Cum. f. below +1/2 f | T-scores (Read from chart) |
|---|---|---|---|---|
| 100 | 1 | 10 | 9.50 | 67 |
| 90 | 2 | 9 | 8.00 | 58 |
| 85 | 3 | 7 | 5.50 | 51 |
| 80 | 2 | 4 | 3.00 | 45 |
| 75 | 1 | 2 | 1.50 | 40 |
| 70 | 1 | 1 | .50 | 33 |

## TABLE I

## A TABLE FOR CALCULATING T-SCORES
(N > 5 < 25)

| Cum.f.. +1/2f | N | | | | | | | | | | | | | | | Cum.f. +1/2f |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | |
| 23.5 | 70 | | | | | | | | | | | | | | | |
| 23 | 68 | | | | | | | | | | | | | | | |
| 22.5 | 66 | 70 | | | | | | | | | | | | | | |
| 22 | 64 | 67 | | | | | | | | | | | | | | |
| 21.5 | 63 | 65 | 70 | | | | | | | | | | | | | |
| 21 | 62 | 63 | 67 | | | | | | | | | | | | | |
| 20.5 | 61 | 62 | 65 | 70 | | | | | | | | | | | | |
| 20 | 60 | 61 | 63 | 67 | | | | | | | | | | | | |
| 19.5 | 59 | 60 | 62 | 65 | 70 | | | | | | | | | | | |
| 19 | 58 | 59 | 61 | 63 | 67 | | | | | | | | | | | |
| 18.5 | 57 | 58 | 60 | 62 | 64 | 70 | | | | | | | | | | |
| 18 | 57 | 58 | 59 | 61 | 63 | 66 | | | | | | | | | | |
| 17.5 | 56 | 57 | 58 | 60 | 62 | 64 | 69 | | | | | | | | | |
| 17 | 55 | 57 | 57 | 59 | 60 | 63 | 66 | | | | | | | | | |
| 16.5 | 55 | 56 | 57 | 58 | 59 | 61 | 64 | 69 | | | | | | | | |
| 16 | 54 | 55 | 56 | 57 | 58 | 60 | 62 | 66 | | | | | | | | |
| 15.5 | 54 | 55 | 55 | 57 | 58 | 59 | 61 | 64 | 68 | | | | | | | |
| 15 | 53 | 54 | 55 | 56 | 57 | 58 | 60 | 62 | 65 | | | | | | | |
| 14.5 | 53 | 54 | 54 | 55 | 56 | 57 | 59 | 61 | 63 | 68 | | | | | | |
| 14 | 52 | 53 | 54 | 54 | 55 | 56 | 58 | 59 | 61 | 65 | | | | | | |
| 13.5 | 52 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 60 | 63 | 68 | | | | | |
| 13 | 51 | 52 | 52 | 53 | 54 | 55 | 56 | 57 | 59 | 61 | 65 | | | | | |
| 12.5 | 51 | 51 | 52 | 52 | 53 | 54 | 55 | 56 | 58 | 60 | 62 | 68 | | | | |
| 12 | 50 | 51 | 51 | 52 | 53 | 53 | 54 | 55 | 57 | 59 | 61 | 64 | | | | |
| 11.5 | 49 | 50 | 51 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 59 | 62 | 67 | | | |
| 11 | 49 | 49 | 50 | 51 | 51 | 52 | 53 | 54 | 55 | 56 | 58 | 60 | 64 | | | |
| 10.5 | 48 | 49 | 49 | 50 | 51 | 51 | 52 | 53 | 54 | 55 | 57 | 59 | 61 | 67 | | |
| 10 | 48 | 48 | 49 | 49 | 50 | 51 | 51 | 52 | 53 | 54 | 56 | 57 | 60 | 63 | | |
| 9.5 | 47 | 48 | 48 | 49 | 49 | 50 | 51 | 51 | 52 | 53 | 55 | 56 | 58 | 61 | 67 | 9.5 |
| 9 | 47 | 47 | 48 | 48 | 49 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 57 | 59 | 63 | 9 |
| 8.5 | 46 | 46 | 47 | 48 | 48 | 49 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 58 | 60 | 8.5 |
| 8 | 46 | 46 | 46 | 47 | 47 | 48 | 49 | 49 | 50 | 51 | 52 | 53 | 54 | 56 | 58 | 8 |
| 7.5 | 45 | 45 | 46 | 46 | 47 | 47 | 48 | 49 | 49 | 50 | 51 | 52 | 53 | 54 | 57 | 7.5 |
| 7 | 45 | 45 | 45 | 46 | 46 | 47 | 47 | 48 | 48 | 49 | 50 | 51 | 52 | 53 | 55 | 7 |
| 6.5 | 44 | 44 | 45 | 45 | 45 | 46 | 46 | 47 | 48 | 48 | 49 | 50 | 51 | 52 | 54 | 6.5 |
| 6 | 43 | 43 | 44 | 44 | 45 | 45 | 46 | 46 | 47 | 47 | 48 | 49 | 50 | 51 | 53 | 6 |
| 5.5 | 43 | 43 | 43 | 43 | 44 | 44 | 45 | 45 | 46 | 47 | 47 | 48 | 49 | 50 | 51 | 5.5 |
| 5 | 42 | 42 | 43 | 43 | 43 | 44 | 44 | 45 | 45 | 46 | 46 | 47 | 48 | 49 | 50 | 5 |
| 4.5 | 41 | 42 | 42 | 42 | 42 | 43 | 43 | 44 | 44 | 45 | 45 | 46 | 47 | 48 | 49 | 4.5 |
| 4 | 40 | 41 | 41 | 41 | 42 | 42 | 42 | 43 | 43 | 44 | 44 | 45 | 46 | 47 | 47 | 4 |
| 3.5 | 39 | 40 | 40 | 40 | 41 | 41 | 41 | 42 | 42 | 43 | 43 | 44 | 45 | 46 | 46 | 3.5 |
| 3 | 38 | 39 | 39 | 39 | 40 | 40 | 40 | 41 | 41 | 41 | 42 | 43 | 43 | 44 | 45 | 3 |
| 2.5 | 37 | 38 | 38 | 38 | 38 | 39 | 39 | 39 | 40 | 40 | 41 | 41 | 42 | 42 | 43 | 2.5 |
| 2 | 36 | 37 | 37 | 37 | 37 | 37 | 38 | 38 | 39 | 39 | 39 | 40 | 40 | 41 | 42 | 2 |
| 1.5 | 34 | 35 | 35 | 35 | 36 | 36 | 36 | 36 | 37 | 37 | 38 | 38 | 39 | 39 | 40 | 1.5 |
| 1 | 32 | 33 | 33 | 33 | 33 | 34 | 34 | 34 | 35 | 35 | 35 | 36 | 36 | 37 | 37 | 1 |
| .5 | 30 | 30 | 30 | 30 | 30 | 30 | 31 | 31 | 32 | 32 | 32 | 32 | 33 | 33 | 33 | .5 |

| Cum.f. +1/2f | N | | | | Cum.f. +1/2f |
|---|---|---|---|---|---|
| | 6 | 7 | 8 | 9 | |
| 0.5 | 36 | 35 | 35 | 34 | 0.5 |
| 1.0 | 40 | 39 | 39 | 38 | 1.0 |
| 1.5 | 43 | 42 | 41 | 40 | 1.5 |
| 2.0 | 46 | 44 | 43 | 42 | 2.0 |
| 2.5 | 48 | 46 | 45 | 44 | 2.5 |
| 3.0 | 50 | 48 | 47 | 46 | 3.0 |
| 3.5 | 52 | 50 | 49 | 47 | 3.5 |
| 4.0 | 54 | 52 | 50 | 49 | 4.0 |
| 4.5 | 57 | 54 | 51 | 50 | 4.5 |
| 5.0 | 60 | 56 | 53 | 51 | 5.0 |
| 5.5 | 64 | 58 | 55 | 53 | 5.5 |
| | | 61 | 57 | 54 | 6.0 |
| | | 65 | 59 | 56 | 6.5 |
| | | | 61 | 58 | 7.0 |
| | | | 65 | 60 | 7.5 |
| | | | | 62 | 8.0 |
| | | | | 66 | 8.5 |

## TABLE II

## A TABLE FOR CALCULATING T-SCORES
## $(N > 24 < 51)$

| Cum.f. +1/2f | 50 | 49 | 48 | 47 | 46 | 45 | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 34 | 33 | 32 | 31 | 30 | 29 | 28 | 27 | 26 | 25 | Cum.f. +1/2f |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49.5 | 73 | | | | | | | | | | | | | | | | | | | | | | | | | | 49.5 |
| 49 | 71 | | | | | | | | | | | | | | | | | | | | | | | | | | 49 |
| 48.5 | 69 | 73 | | | | | | | | | | | | | | | | | | | | | | | | | 48.5 |
| 48 | 68 | 70 | | | | | | | | | | | | | | | | | | | | | | | | | 48 |
| 47.5 | 67 | 68 | 73 | | | | | | | | | | | | | | | | | | | | | | | | 47.5 |
| 47 | 66 | 67 | 70 | | | | | | | | | | | | | | | | | | | | | | | | 47 |
| 46.5 | 65 | 66 | 68 | 73 | | | | | | | | | | | | | | | | | | | | | | | 46.5 |
| 46 | 64 | 65 | 67 | 70 | | | | | | | | | | | | | | | | | | | | | | | 46 |
| 45.5 | 63 | 65 | 66 | 68 | 73 | | | | | | | | | | | | | | | | | | | | | | 45.5 |
| 45 | 63 | 64 | 65 | 67 | 70 | | | | | | | | | | | | | | | | | | | | | | 45 |
| 44.5 | 62 | 63 | 65 | 66 | 68 | 73 | | | | | | | | | | | | | | | | | | | | | 44.5 |
| 44 | 62 | 63 | 64 | 65 | 67 | 70 | | | | | | | | | | | | | | | | | | | | | 44 |
| 43.5 | 61 | 62 | 63 | 64 | 66 | 68 | 73 | | | | | | | | | | | | | | | | | | | | 43.5 |
| 43 | 61 | 62 | 63 | 64 | 65 | 67 | 70 | | | | | | | | | | | | | | | | | | | | 43 |
| 42.5 | 60 | 61 | 62 | 63 | 64 | 66 | 68 | 73 | | | | | | | | | | | | | | | | | | | 42.5 |
| 42 | 60 | 61 | 62 | 62 | 64 | 65 | 67 | 70 | | | | | | | | | | | | | | | | | | | 42 |
| 41.5 | 59 | 60 | 61 | 62 | 63 | 64 | 66 | 68 | 73 | | | | | | | | | | | | | | | | | | 41.5 |
| 41 | 59 | 60 | 61 | 61 | 62 | 63 | 65 | 67 | 70 | | | | | | | | | | | | | | | | | | 41 |
| 40.5 | 59 | 59 | 60 | 61 | 62 | 63 | 64 | 66 | 68 | 73 | | | | | | | | | | | | | | | | | 40.5 |
| 40 | 58 | 59 | 60 | 60 | 61 | 62 | 63 | 65 | 67 | 70 | | | | | | | | | | | | | | | | | 40 |
| 39.5 | 58 | 59 | 59 | 60 | 61 | 62 | 63 | 64 | 66 | 68 | 73 | | | | | | | | | | | | | | | | 39.5 |
| 39 | 58 | 58 | 59 | 60 | 60 | 61 | 62 | 63 | 65 | 67 | 70 | | | | | | | | | | | | | | | | 39 |
| 38.5 | 58 | 58 | 59 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 68 | 73 | | | | | | | | | | | | | | | 38.5 |
| 38 | 57 | 58 | 58 | 59 | 59 | 60 | 61 | 62 | 63 | 65 | 67 | 70 | | | | | | | | | | | | | | | 38 |
| 37.5 | 57 | 57 | 58 | 58 | 59 | 60 | 61 | 61 | 62 | 64 | 66 | 68 | 72 | | | | | | | | | | | | | | 37.5 |
| 37 | 57 | 57 | 57 | 58 | 59 | 59 | 60 | 61 | 62 | 63 | 65 | 67 | 69 | | | | | | | | | | | | | | 37 |
| 36.5 | 56 | 57 | 57 | 58 | 58 | 59 | 60 | 60 | 61 | 62 | 64 | 66 | 68 | 72 | | | | | | | | | | | | | 36.5 |
| 36 | 56 | 56 | 57 | 57 | 58 | 58 | 59 | 60 | 61 | 62 | 63 | 65 | 66 | 69 | | | | | | | | | | | | | 36 |
| 35.5 | 56 | 56 | 56 | 57 | 58 | 58 | 59 | 59 | 60 | 61 | 62 | 64 | 65 | 67 | 72 | | | | | | | | | | | | 35.5 |
| 35 | 55 | 56 | 56 | 57 | 57 | 58 | 58 | 59 | 60 | 60 | 62 | 63 | 64 | 66 | 69 | | | | | | | | | | | | 35 |
| 34.5 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 58 | 57 | 60 | 61 | 62 | 63 | 65 | 67 | 72 | | | | | | | | | | | 34.5 |
| 34 | 55 | 55 | 56 | 56 | 57 | 57 | 57 | 58 | 59 | 59 | 60 | 61 | 63 | 64 | 65 | 69 | | | | | | | | | | | 34 |
| 33.5 | 54 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 58 | 59 | 60 | 61 | 62 | 63 | 65 | 67 | 72 | | | | | | | | | | 33.5 |
| 33 | 54 | 55 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 58 | 59 | 60 | 61 | 62 | 64 | 66 | 69 | | | | | | | | | | 33 |
| 32.5 | 54 | 54 | 55 | 55 | 56 | 56 | 58 | 57 | 58 | 58 | 59 | 60 | 60 | 62 | 63 | 65 | 67 | 72 | | | | | | | | | 32.5 |
| 32 | 54 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 58 | 59 | 60 | 61 | 62 | 64 | 66 | 69 | | | | | | | | | 32 |
| 31.5 | 53 | 54 | 54 | 55 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 59 | 59 | 60 | 61 | 63 | 65 | 67 | 72 | | | | | | | | 31.5 |
| 31 | 53 | 53 | 54 | 54 | 55 | 55 | 55 | 56 | 56 | 57 | 58 | 58 | 59 | 60 | 61 | 62 | 64 | 66 | 69 | | | | | | | | 31 |
| 30.5 | 53 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 56 | 57 | 58 | 58 | 59 | 60 | 61 | 63 | 65 | 67 | 72 | | | | | | | 30.5 |
| 30 | 53 | 53 | 53 | 54 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 59 | 60 | 60 | 62 | 64 | 65 | 69 | | | | | | | 30 |
| 29.5 | 52 | 53 | 53 | 53 | 54 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 58 | 58 | 59 | 60 | 61 | 63 | 64 | 67 | 71 | | | | | | 29.5 |
| 29 | 52 | 52 | 53 | 53 | 53 | 54 | 54 | 55 | 55 | 55 | 56 | 57 | 57 | 58 | 59 | 59 | 60 | 62 | 63 | 65 | 68 | | | | | | 29 |
| 28.5 | 52 | 52 | 52 | 53 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 59 | 60 | 61 | 62 | 64 | 66 | 71 | | | | | 28.5 |
| 28 | 52 | 52 | 52 | 52 | 53 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 58 | 58 | 59 | 60 | 61 | 63 | 65 | 68 | | | | | 28 |
| 27.5 | 51 | 52 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 54 | 55 | 56 | 56 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 64 | 66 | 71 | | | | 27.5 |
| 27 | 51 | 51 | 52 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 57 | 58 | 59 | 60 | 61 | 63 | 65 | 68 | | | | 27 |
| 26.5 | 51 | 51 | 51 | 52 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 58 | 59 | 59 | 60 | 62 | 64 | 66 | 71 | | | 26.5 |
| 26 | 51 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 58 | 59 | 60 | 61 | 63 | 65 | 68 | | | 26 |
| 25.5 | 50 | 51 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 54 | 55 | 56 | 56 | 57 | 58 | 58 | 59 | 60 | 62 | 64 | 66 | 71 | | 25.5 |
| 25 | 50 | 50 | 51 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 58 | 59 | 60 | 61 | 63 | 64 | 68 | | 25 |
| 24.5 | 50 | 50 | 50 | 51 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 57 | 57 | 58 | 59 | 60 | 62 | 63 | 66 | 70 | 24.5 |
| 24 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 51 | 52 | 52 | 53 | 53 | 53 | 54 | 54 | 55 | 56 | 56 | 57 | 58 | 58 | 59 | 61 | 62 | 64 | 68 | 24 |
| 23.5 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 51 | 52 | 52 | 53 | 53 | 53 | 54 | 54 | 55 | 56 | 56 | 57 | 58 | 59 | 60 | 61 | 63 | 66 | 23.5 |
| 23 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 57 | 57 | 58 | 59 | 60 | 62 | 64 | 23 |
| 22.5 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 63 | 22.5 |
| 22 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 56 | 56 | 57 | 58 | 59 | 60 | 62 | 22 |
| 21.5 | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 56 | 57 | 57 | 58 | 59 | 61 | 21.5 |
| 21 | 48 | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 53 | 53 | 53 | 54 | 55 | 55 | 56 | 57 | 58 | 59 | 60 | 21 |
| 20.5 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 56 | 56 | 57 | 58 | 59 | 20.5 |
| 20 | 47 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 56 | 56 | 57 | 58 | 20 |
| 19.5 | 47 | 47 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 55 | 55 | 56 | 57 | 58 | 19.5 |
| 19 | 47 | 47 | 47 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 55 | 55 | 56 | 57 | 19 |
| 18.5 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 52 | 52 | 53 | 54 | 54 | 55 | 56 | 56 | 18.5 |
| 18 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 56 | 18 |
| 17.5 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 54 | 55 | 17.5 |
| 17 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 55 | 17 |
| 16.5 | 46 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 48 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 52 | 53 | 53 | 54 | 16.5 |
| 16 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 52 | 53 | 54 | 16 |
| 15.5 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 52 | 53 | 15.5 |
| 15 | 45 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 49 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 53 | 15 |
| 14.5 | 44 | 45 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 48 | 48 | 48 | 48 | 49 | 49 | 50 | 50 | 50 | 51 | 51 | 52 | 14.5 |
| 14 | 44 | 44 | 44 | 45 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 49 | 49 | 50 | 50 | 50 | 51 | 52 | 14 |
| 13.5 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 48 | 49 | 49 | 50 | 50 | 50 | 51 | 13.5 |
| 13 | 43 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 48 | 48 | 48 | 49 | 49 | 50 | 50 | 51 | 13 |
| 12.5 | 43 | 43 | 44 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 48 | 48 | 48 | 49 | 49 | 50 | 50 | 12.5 |
| 12 | 43 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 47 | 47 | 47 | 47 | 48 | 48 | 49 | 49 | 49 | 12 |
| 11.5 | 42 | 43 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 47 | 48 | 48 | 49 | 49 | 11.5 |
| 11 | 42 | 42 | 43 | 43 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 45 | 46 | 46 | 46 | 47 | 47 | 47 | 48 | 48 | 48 | 11 |
| 10.5 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 46 | 46 | 46 | 46 | 47 | 47 | 48 | 48 | 10.5 |
| 10 | 42 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 46 | 46 | 46 | 47 | 47 | 47 | 10 |
| 9.5 | 41 | 41 | 41 | 42 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 45 | 46 | 46 | 47 | 47 | 9.5 |
| 9 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 44 | 45 | 45 | 45 | 46 | 46 | 46 | 9 |
| 8.5 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 48 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 44 | 45 | 45 | 46 | 46 | 8.5 |
| 8 | 40 | 40 | 40 | 40 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 45 | 45 | 45 | 8 |
| 7.5 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 43 | 44 | 44 | 44 | 45 | 7.5 |
| 7 | 39 | 39 | 39 | 40 | 40 | 40 | 40 | 40 | 40 | 41 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 44 | 44 | 44 | 7 |
| 6.5 | 39 | 39 | 39 | 39 | 39 | 39 | 39 | 40 | 40 | 40 | 40 | 40 | 41 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 42 | 43 | 43 | 43 | 44 | 6.5 |
| 6 | 38 | 38 | 38 | 39 | 39 | 39 | 39 | 39 | 39 | 40 | 40 | 40 | 40 | 40 | 40 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 42 | 43 | 43 | 6 |
| 5.5 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 39 | 39 | 39 | 39 | 39 | 40 | 40 | 40 | 40 | 40 | 40 | 41 | 41 | 41 | 41 | 41 | 42 | 42 | 42 | 5.5 |
| 5 | 37 | 37 | 37 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 39 | 39 | 39 | 39 | 40 | 40 | 40 | 40 | 40 | 40 | 41 | 41 | 41 | 41 | 42 | 5 |
| 4.5 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 38 | 38 | 38 | 38 | 38 | 38 | 39 | 39 | 39 | 39 | 39 | 40 | 40 | 40 | 40 | 40 | 41 | 41 | 4.5 |
| 4 | 36 | 36 | 36 | 36 | 36 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 38 | 38 | 38 | 38 | 38 | 39 | 39 | 39 | 39 | 39 | 40 | 40 | 40 | 4 |
| 3.5 | 35 | 35 | 35 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 36 | 37 | 37 | 37 | 37 | 37 | 37 | 38 | 38 | 38 | 38 | 38 | 39 | 39 | 39 | 3.5 |
| 3 | 34 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 36 | 36 | 36 | 36 | 36 | 36 | 37 | 37 | 37 | 37 | 37 | 38 | 38 | 38 | 3 |
| 2.5 | 33 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 35 | 35 | 35 | 35 | 35 | 35 | 36 | 36 | 36 | 36 | 36 | 37 | 37 | 37 | 2.5 |
| 2 | 32 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 34 | 34 | 34 | 34 | 34 | 34 | 35 | 35 | 35 | 35 | 35 | 36 | 36 | 36 | 2 |
| 1.5 | 31 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 34 | 34 | 34 | 34 | 34 | 34 | 1.5 |
| 1 | 29 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 31 | 32 | 32 | 32 | 32 | 32 | 32 | 1 |
| .5 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 29 | 29 | 29 | 29 | 29 | 30 | .5 |

Note: N is encircled figure.

This table should prove to be of value to college instructors and classroom teachers who desire to normalize distributions of scores of classroom tests and report scores as normalized T-scores. This table may be extended indefinitely, but it is felt that the maximum N of 50 will suffice for the great majority of classroom situations.

## REFERENCES

1. Garrett, H. E. Statistics in Psychology and Education (New York: Longmans, Green and Co., 1954), Table G, p. 433.
2. Edwards, A. L. Statistical Methods in the Behavioral Sciences (New York: Rinehart and Co., 1954), Table XI, p. 511.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# ESTIMATION OF TEST RELIABILITY BY ANALYSIS OF VARIANCE TECHNIQUE*

CYRIL J. HOYT and P. R. KRISHNAIAH
University of Minnesota

## Introduction

IN AN EARLIER paper, Hoyt[1] showed a method of determining the best linear estimate of the error variance appropriate for expressing an internal consistency reliability coefficient for a total test score obtained by adding the scores on items scored either one or zero. In this derivation it was assumed that the pq errors of measurement in the responses of p students to each of the q items were independent. Later investigation showed that the procedure described in this earlier paper is entirely appropriate when:

1. items are scored in more than two categories[2],
2. "items" are ratings[3], test scores[4] or other measures[5].

Further theoretical development in the area of components of variance estimation has enabled the authors to show that certain assumptions regarding the independence of the errors of measurement and of independence of main effects (individuals and items) are not necessary in order that Hoyt's 1941 formulation of the reliability coefficient estimate yield appropriate ratios of expected values of the mean squares.

When item effects are assumed to be intercorrelated (see models 5 and 6) the expected values of the two variance components (A) and (C) used in estimating the reliability coefficient are precisely equal to those obtained for models 1, 2, 3, and 4 in which independence is assumed as in the 1941 derivation.

When errors of measurement are assumed to be correlated for the responses of a given individual on different items (see models 7, 8, and 9) the expected values of both variance components A (among individuals) and C (error) are different from those obtained for the first six models. However, the ratio of (A-C) to A still provides a suitable estimate of the internal consistency reliability coefficients provided the user recognizes what he has.

The definition of fixed, mixed and random models follows:

The model equation for each of the nine lines of Table I is $X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ where

$$i = 1, 2.\dots p$$
$$j = 1, 2.\dots q$$

where $\mu$, $\alpha_i$, $\beta_j$ are the general effect, $i^{th}$ individual effect, and $j^{th}$ item effect. $\epsilon_{ij}$'s are errors which are normally distributed. This equation is suitable for all three types of models with different assumptions on the $\alpha$'s and $\beta$'s. We may suppose that the individuals and the items are drawn randomly from their corresponding populations which contain P individuals and Q items respectively.

### A. Fixed Model

When $P = p$, and $Q = q$, then the above equation represents a fixed model. In this case, the only random element is the error term $\epsilon_{ij}$. When describing the reliability coefficient for a given test form administered to a given group of individuals the fixed model is appropriate.

### B. Mixed Model

When $P = p$ and $Q \gg q$ or when $P \gg p$ and $Q = q$, then the equation represents a mixed model. (The symbol $\gg$ is used to designate the relation as much greater.) This model is suitable for considering the reliability obtained as an indicator of that appropriate for either measuring a given group of people on a domain of items represented by the sample of q items or, alternatively, for generalizing to a population of P subjects for any given administration of a certain test of q items.

### C. Random Model

When $P \gg p$ and $Q \gg q$ then the above equation represents a random model. This model is suitable for estimating the reliability coefficient for

---

*All footnotes will be found at end of article.

TABLE I

EXPECTED VALUES OF MEAN SQUARES FOR NINE MODELS

| Model | (A) Among Individuals | (B) Among Items | (C) Error | A - C |
|---|---|---|---|---|
| 1. Fixed model<br>Individual item effects fixed<br>Errors uncorrelated | $\sigma^2 + \frac{q}{p-1}\Sigma\alpha_i^2$ | $\sigma^2 + \frac{p}{q-1}\Sigma\beta_j^2$ | $\sigma^2$ | $\frac{q}{p-1}\Sigma\alpha_i^2$ |
| 2. Mixed model<br>Errors uncorrelated<br>Item effect random<br>Individual effects fixed | $\sigma^2 + \frac{q}{p-1}\Sigma\alpha_i^2$ | $\sigma^2 + p\sigma_\beta^2$ | $\sigma^2$ | $\frac{q}{p-1}\Sigma\alpha_i^2$ |
| 3. Mixed model<br>Errors uncorrelated<br>Item effects fixed<br>Individual effects random | $\sigma^2 + q\sigma_\alpha^2$ | $\sigma^2 + \frac{p}{q-1}\Sigma\beta_j^2$ | $\sigma^2$ | $q\sigma_\alpha^2$ |
| 4. Random model<br>Errors uncorrelated<br>Item effects random<br>Individual effects random | $\sigma^2 + q\sigma_\alpha^2$ | $\sigma^2 + p\sigma_\beta^2$ | $\sigma^2$ | $q\sigma_\alpha^2$ |
| 5. Random model<br>Errors uncorrelated<br>Item effects correlated<br>Individual effects random | $\sigma^2 + q\sigma_\alpha^2$ | $\sigma^2 + p(1 - \rho)\,\sigma_\beta^2$ | $\sigma^2$ | $q\sigma_\alpha^2$ |
| 6. Mixed model<br>Errors uncorrelated<br>Item effects correlated<br>Individual effects fixed | $\sigma^2 + \frac{q}{p-1}\Sigma\alpha_i^2$ | $\sigma^2 + p(1 - \rho)\,\sigma_\beta^2$ | $\sigma^2$ | $\frac{q}{p-1}\Sigma\alpha_i^2$ |
| 7. Fixed model<br>Errors correlated<br>on different items for<br>same individual | $\sigma^2[1 + (q - 1)\rho] + \frac{q}{p-1}\Sigma\alpha_i^2$ | $\sigma^2(1 - \rho) + \frac{p}{q-1}\Sigma\beta_j^2$ | $\sigma^2(1 - \rho)$ | $\frac{q}{p-1}\Sigma\alpha_i^2 + \rho q\sigma^2$ |
| 8. Mixed model<br>Errors correlated as in 7<br>Individual effects fixed<br>Item effects random | $\sigma^2[1 + (q - 1)\rho] + \frac{q}{p-1}\Sigma\alpha_i^2$ | $\sigma^2(1 - \rho) + p\sigma_\beta^2$ | $\sigma^2(1 - \rho)$ | $\frac{q}{p-1}\Sigma\alpha_i^2 + \rho q\sigma^2$ |
| 9. Random model<br>Errors correlated as in 7<br>Item effects random<br>Individual effects random | $\sigma^2[1 + (q - 1)\rho] + q\sigma_\alpha^2$ | $\sigma^2(1 - \rho) + p\sigma_\beta^2$ | $\sigma^2(1 - \rho)$ | $q\sigma_\alpha^2 + \rho q\sigma^2$ |

an infinite population of subjects or a finite one o size P, on the whole domain of Q items.

Columns 2, 3, and 4 of Table I show the expected values of the mean squares components in nin different models that are each specified by the assumed characteristics listed in column one. Thes nine models are all useful for specifying interna consistency reliability coefficients if $X_{ij}$ is considered the score earned by i - th individual's response to the j - th item on an examination. I these nine models the interaction term is used a error as is usually done in cases providing onl one observation per cell. By making different assumptions about the parameters in the equation th nine different models each described in successive lines of Table I arise. There are a numbe of other models which seem useful and for whic we have not yet studied the variance componen and reliability estimates.

The first six entries in the last column of Table I are the expected values of the mean square of the true scores or the numerator of the reliability coefficient appropriate for each line of th table. The denominators of the reliability coeff cients in every case are the entries in column In the models specified in the last three lines, th entries in the last column seem to contain mo than the variance of the true scores. Henc though these could be defined as the numerators the reliability coefficients, they do not appear be appropriate estimates for the variance of t true scores.

Let us consider now the first four models which all the effects represented by the terms the equation are independent. In these cases, c notes that the usual procedure for estimating t internal consistency reliability coefficient giv adequate estimates of the ratio of the variance the true scores to the variance of the obtain scores. The last column of Table I gives only numerator of the reliability coefficient obtained taking one minus the ratio of the error mean squa to the among students' mean square. The deno inator of the reliability coefficient is (A) of corresponding row or the mean square for amo individuals. Rows 5 and 6 show expected valu for random and mixed models, respectively, which the item effects are correlated with one other to the same extent ($\rho$) for each pair of items. The column indicating the expected value of the "among items" mean square gives an expression involving this common correlation $\rho$ between items. However, as long as the errors of measurement in the several items are considered independent of each other and independent of the item and individual effects, the usual mean square ratio for estimating the coefficient of reliability gives a numerator which is an appropriate estimate of the variance of the true scores.

In models 7, 8, and 9, however, the numerator of the usual estimate of the reliability coefficient contains a term $\rho q \sigma^2$ which is not zero unless $\rho$ is zero. In these three models, it is assumed that the errors of measurement on pairs of items are correlated with a common co-variance for each individual.

## FOOTNOTES

* Work on this paper was partially supported by the Bureau of Educational Research, University of Minnesota.

1. Cyril J. Hoyt, "Test Reliability Estimated by Analysis of Variance," Psychometrika, VI (1941), pp. 153-60.

2. Cyril J. Hoyt and Clayton L. Stunkard, "Estimation of Test Reliability for Unrestricte Item Scoring Methods," Educational and Psy chological Measurement, XII (1951), pp 756-58.

3. Robert L. Ebel, "Estimation of the Reliabilit of Ratings," Psychometrika, XVI (1951 pp. 407-24.

4. P. Horst, "A Generalized Expression for th Reliability of Measures," Psychometrik XIV (1949), pp. 21-31.

5. Howard W. Alexander, "The Estimation o Reliability When Several Trials are Avail able," Psychometrika, XII (1947), pp. 79-9

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# ON TEACHING F INSTEAD OF t

NORMAN H. ANDERSON
Yale University*

IT IS WELL known that the square of a t ratio on n df is interpretable as an F ratio on 1 and n df. For the case of one or two means then, the values of t and F are equivalents, and lead to identical probability and inferential statements. Since F is the more general, applying as well to the case of more than two means, one might well suspect that the t test could be retired from use. Because the t test constitutes such a large part of current statistical pedagogy and practice, it may be useful to consider the matter in more detail. The main considerations which seem to the writer to be relevant are summarized briefly below.[1] These considerations are applicable principally to the first few statistics courses in the applied sciences and have less relevance to more theoretical courses.

In favor of t, it can be argued that it has greater intuitive appeal. To measure the difference between two means against its standard error may be simpler to comprehend and result in easier learning than does the variance partitioning technique of F. However, as soon as it is admitted that t is insufficient to meet everyday needs, so that F must be taught anyway, this argument loses its power. It may also be suggested that whatever greater immediacy of meaning may attach to t is not entirely its natural due, but rather a product of the present organization of statistics courses.

The t test may also be preferred because of its relation to some of the problems arising in estimation, such as the construction of confidence intervals. Algebraically, of course, it makes little difference whether one proceeds from t or from F. Thus, with F, the 95 percent confidence interval is obtained using $\sqrt{s_m^2 F(.95; 1, n)}$. The interpretation of the confidence interval could then be taken on faith (and even serve usefully as the basis for the significance test for the case of one or two means). However, it must be admitted that the confidence interval idea seems easier to comprehend when it is based on the standard picture of the symmetrical t distribution. This possible advantage of t must be measured against its shortcomings in problems of significance testing which are discussed below.

A further point in favor of t stems from a danger of the analysis of variance. Ordinarily, when pair-wise comparisons among several conditions are to be made, the error mean square from the overall F analysis is used throughout. When the condition variances are reasonably equal, this procedure gives increased sensitivity since the error is estimated more precisely. If there is marked heterogeneity of variance, however, such tests may be biased, and it may be preferable to test each pair of conditions using only the data from that pair (although this does not entirely escape the difficulty). Of course, the two condition F could be used here as easily as t. Nevertheless, it must be admitted that routine application of the analysis of variance does facilitate overlooking the problem.

Against these arguments must be set a number of negative aspects of t. For situations with just one or two conditions, of course, it is a matter of statistical indifference whether t or F is used. Either yields the same inferences, as noted above. Moreover, since $t^2 = F$, there need be little difference in computational efficiency. It may be noted, however, that the standard computing rules for t, in contrast to those for F, are not efficient and are thus more liable to arithmetical error.

When more than two conditions are involved, piece-meal t testing presents certain dangers.[2] The one commonly emphasized is the increased risk of thereby obtaining "significant" differences which the data do not in fact support. Curiously enough, the t test approach also increases the likelihood of the obverse error, namely, failing to detect results which the data are quite adequate to declare real. This can happen in a number of ways, which trend tests may serve to illustrate. In investigations in which one of the independent variables is numerical, the between conditions variation can often be expected on prior grounds to reside largely in the linear, or linear + quadratic, component of the curve. Pair-wise comparisons of means may fail to reveal even a fairly strong real effect which a trend analysis F will elucidate with ease.

A more fundamental objection to the t test approach concerns the original planning of the investigation. The researcher who thinks in terms of t is less likely to use as incisive and informative a design as can be had for the allotted time and effort. Even when the basic plan is sound, a poor superstructure can, and not infrequently does, prejudice the analysis and interpretation, or entail considerable extra computational cost.

---

*Footnotes will be found at end of article.

Fisher's commendation of the use of factorial designs for controlling extraneous variation may be cited in illustration of this last point. Even with only two main experimental conditions, it is often desirable to counterbalance a second variable, not of great interest in itself, in order to obtain added generality. Unless set up so that the two-way analysis can be and is used, however, the counterbalancing may do more harm than good.

In a similar way, it is often possible to use as a second independent variable some individual difference measure which is correlated with the response measure under study. This procedure, essentially a "randomized blocks" factorial design, may generally be expected not only to increase the sensitivity of the experiment, but also to yield information on the organismic variables.

That the F test is a desirable piece of statistical machinery hardly needs even the incidental justification given here. But if F is to be taught at all, then it evidently follows, since F is the more general, that F alone should be taught. To require the mastery of two sets of rules where one will do is an improvident use of the student's time.

For statistics courses in which time is insufficient for both tests, the choice of F would accomplish the same ends as the choice of t. If more time were available, the elimination of t would permit the development of some of the simpler ramifications of the analysis of variance, such as two-way designs, specific comparisons, and trend tests. Moreover, teaching F from the start would set up conditions for maximal positive transfer later in learning since the computing rules for F follow a uniform and elegant logic in all situations. It might also be noted in this connection that the computational procedures for correlation and regression analysis can be subsumed under the analysis of variance format as well.

Finally, it may be questioned whether learning the t test does not produce a fair amount of negative transfer. The evidence available to the writer, admittedly anecdotal in nature, suggests that students who have learned t are reluctant to abandon this skill when necessary. Having a technique which may actually, if not always advisedly, be used, they tend to regard even workaday designs as "fancy" or "not worth the trouble at this stage of investigation." Thus, if they use the design at all, they do so only with pain, and so fail to acquire the familiarity with the new technique that would come were it regarded as just another application of the F test they already knew.

It would thus appear that t is superfluous, since F may always be computed instead; inadequate, since it does not apply as generally as does F; dangerous, since its use may increase errors of inference, as well as lead to weak experiments; and a waste of time, since teaching F will provide more transfer to later learning. On the balance, it seems fair to conclude that the t test should be struck from the curriculum.

More practical experience is needed in order to integrate the analysis of variance with beginning statistics on the conceptual level. However, it is appropriate to sketch briefly the sequence of computational routines which is indicated by the argument of this note.

1. Computation of the sample variance is done using the analysis of variance table:

| Source | df | SS | MS |
|---|---|---|---|
| Mean | 1 | $T^2/N$ | |
| Error (Variability) | N-1 | $\Sigma X^2 - T^2/N$ | |

The entry, $MS_{error}$, then furnishes an unbiased estimate of the population variance. The scaffolding provided by the tabular format is unnecessary, of course, but seems to serve a useful mnemonic function. Its real benefit accrues in Step 2.

2. The test of significance for a single sample is most appropriately carried out on difference scores, or on other scores which have a mean of 0 under the null hypothesis. A single division appended to the Step 1 routine yields the relevant F ratio.

3. With two samples of difference scores, the analysis of variance table is extended one step further:

| Source | df | SS | MS | F |
|---|---|---|---|---|
| Mean | 1 | | | |
| Between Groups | 1 | | | |
| Error | N-2 | | | |

4. It is straightforward to extend Step 3 to the case of several samples and it seems advisable to do so promptly in order to avoid the development of a superstition that the technique only works for one or two. The routine remains the same when the scores are arbitrary, rather than difference scores. It is only necessary to observe that the F for Mean is no longer meaningful or important and may be omitted from the table.

5. A simple extension of the computing routine yields the calculation of r for a single sample:

| Source | XX | XY | YY |
|---|---|---|---|
| Mean | $T_X^2/N$ | $T_X T_Y/N$ | $T_Y^2/N$ |
| Error | $\Sigma X^2 - T_X^2/N$ | $\Sigma XY - T_X T_Y/N$ | $\Sigma Y^2 - T_Y^2/N$ |

The square of the correlation coefficient is estimated from the entries in the Error line, $r^2 = SS_{XY}^2/SS_{XX}SS_{YY}$. The test for significance may be made by consulting a table for r or, with an additional calculation, by using F. When the calculation of r is taught this way, there is less likelihood of confusedly agglomerating the between and within samples regressions when more than a single sample is involved.

It is suggested that $\Sigma X^2$ be given for most of the homework problems. This will greatly decrease the time spent on each problem and, by thus allowing a greater variety of problems to be assigned, increase the learning efficiency. Whether or not further applications of the analysis of variance are given in the first course will depend on the available time, and on whether the course is oriented toward experimental or observational research.

## FOOTNOTES

* Now at the University of California, Los Angeles.

1. For the purposes of this note, the terms "F test" and "analysis of variance" will be taken as synonymous although this is not strictly correct. It will be necessary to allude to a number of uses of the analysis of variance, but these uses are well known and will not be documented.

2. A specific comparison based on some reasonably firm rational or empirical ground may always be tested, of course. However, unless it is based on a single degree of freedom, it will require an F ratio in its own right. It should also be specifically noted that multiple comparison procedures will be superior to F (and t) in certain situations.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# COMPUTATION OF VARIANCE ACCOUNTED FOR IN MULTIPLE CORRELATION

CLINTON I. CHASE
Idaho State College
Pocatello, Idaho

WHEN MULTIPLE correlation techniques are used in the analysis of data, the investigator usually wishes not only to discover a combination of variables which will yield maximum prediction of a criterion, but also he wishes to determine the amount of variance in the criterion which is associated with each of the independent variables. For this purpose Engelhart (1) has submitted the formula:

$$R^2 = \beta_{01}^2 + \beta_{02}^2 + \ldots \beta_{0n}^2 + 2\beta_{01}\beta_{02}r_{12} + 2\beta_{01}\beta_{03}r_{13} + \ldots 2\beta_{0,n-1}\beta_{0n}r_{n-1,n} \qquad [1]$$

Following this procedure, the ith beta squared will produce the variance directly associated with the ith variable. But since the ith variable is usually correlated with other independent variables, it also enters into variance which is shared by several of these other independent variables. The total amount of this shared variance is given by the summation of all of the terms of the form $2\beta_{0i}\beta_{0j}r_{ij}$ which appear in formula [1]. Then, according to Engelhart's procedure, the shared variance is divided among each of the independent variables in the same proportions as the direct variance. The total amount of variance associated with the ith variable, then, would be the ith beta squared, plus a proportional amount of the total shared variance.

Since Engelhart's procedure for partitioning the shared variance is admittedly arbitrary, the purpose of this paper is to present an alternate procedure, which, it is believed, is somewhat less arbitrary in its method of apportioning shared variance, since here the basis for partitioning the shared variance is found in mathematical manipulations.

As a necessary part of this presentation a second technique (2:409-10) for computing the variance in the criterion associated with any given independent variable must be noted. Such a technique is provided in the formula:

$$R^2 = \beta_{01}r_{01} + \beta_{02}r_{02} + \beta_{03}r_{03} + \ldots \beta_{0n}r_{0n} \qquad [2]$$

Using this procedure, one may find the total amount of variance associated with the ith variable in the product $\beta_{0i}r_{0i}$. Although formula [2] is simpler to use than formula [1], especially when the number of independent variables is more than two or three, it has the disadvantage of not yielding a break-down between variance which is directly associated with the ith variable, and that portion of the variance shared among the several independent variables which is to be associated with the ith variable.

It therefore seems that some technique is desirable which would retain simplicity and economy of effort, while yielding a maximum amount of information. Such a system is found in the mathematical analysis of formulas [1] and [2] as shown below.

In a set of normal equations for the least-square estimate of the betas for the multiple linear-regression equation, the r's are found as follows:

$$r_{01} = \beta_{01} + r_{12}\beta_{02} + r_{13}\beta_{03} + r_{14}\beta_{04} \qquad [3]$$
$$r_{02} = \beta_{02} + r_{12}\beta_{01} + r_{23}\beta_{03} + r_{24}\beta_{04} \qquad [4]$$
$$r_{03} = \beta_{03} + r_{13}\beta_{01} + r_{23}\beta_{02} + r_{34}\beta_{04} \qquad [5]$$
$$r_{04} = \beta_{04} + r_{14}\beta_{01} + r_{24}\beta_{02} + r_{34}\beta_{03} \qquad [6]$$

Therefore, the right hand portion of [3] can be substituted for $r_{01}$ in formula [2], the right hand side of [4] can be substituted for $r_{02}$ in [2], and so on. The result would be as follows:

$$\begin{aligned} R^2 = \; & \beta_{01}(\beta_{01} + r_{12}\beta_{02} + r_{13}\beta_{03} + r_{14}\beta_{04}) + \\ & \beta_{02}(\beta_{02} + r_{12}\beta_{01} + r_{23}\beta_{03} + r_{24}\beta_{04}) + \\ & \beta_{03}(\beta_{03} + r_{13}\beta_{01} + r_{23}\beta_{02} + r_{34}\beta_{04}) + \\ & \beta_{04}(\beta_{04} + r_{14}\beta_{01} + r_{24}\beta_{02} + r_{34}\beta_{03}) \end{aligned} \qquad [7]$$

Multiplying to eliminate the parentheses one may find the following results:

---

*The author wishes to acknowledge the advice received from R. F. Jarrett of the Department of Psychology, University of California, Berkeley, California.

$$\begin{aligned} R^2 = {} & \beta_{O1}^2 + r_{12}\beta_{O2}\beta_{O1} + r_{13}\beta_{O3}\beta_{O1} + r_{14}\beta_{O4}\beta_{O1} + \\ & \beta_{O2}^2 + r_{12}\beta_{O1}\beta_{O2} + r_{23}\beta_{O3}\beta_{O2} + r_{24}\beta_{O4}\beta_{O2} + \\ & \beta_{O3}^2 + r_{13}\beta_{O1}\beta_{O3} + r_{23}\beta_{O2}\beta_{O3} + r_{34}\beta_{O4}\beta_{O3} + \\ & \beta_{O4}^2 + r_{14}\beta_{O1}\beta_{O4} + r_{24}\beta_{O2}\beta_{O4} + r_{34}\beta_{O3}\beta_{O4} \end{aligned} \qquad [8]$$

It can now be seen that summating like terms in [8] will produce formula [1] as it would be written in a four variable problem:

$$\begin{aligned} R^2 = {} & \beta_{O1}^2 + \beta_{O2}^2 + \beta_{O3}^2 + \beta_{O4}^2 + 2\beta_{O1}\beta_{O2}r_{12} + \\ & 2\beta_{O1}\beta_{O3}r_{13} + 2\beta_{O1}\beta_{O4}r_{14} + 2\beta_{O2}\beta_{O3}r_{23} + \\ & 2\beta_{O2}\beta_{O4}r_{24} + 2\beta_{O3}\beta_{O4}r_{3} \end{aligned} \qquad [9]$$

Therefore, from formula [3] and [7] it is apparent that the combined direct and shared variance associated with variable one, as given by the $\beta_{oi}r_{oi}$ product in formula [2], is actually equal to the beta squared, plus one-half of the summation of all the covariance terms in formula [9], of the form $2\beta_{O1}\beta_{Oj}r_{1j}$, which include the beta of variable one. Likewise, it can be generalized that the total direct and shared variance in the criterion associated with the ith independent variable is given by the square of the beta for the ith variable, plus half of all the covariance terms in formula [1] which include the beta for the ith variable.

It appears, then, from a mathematical viewpoint that the variance which the ith variable shares with other variables will not necessarily be distributed in the same proportion as is the variance directly associated with that variable, as Engelhart proposed. Noting the relationship between formula [1] and formula [2], one may now compute both the variance which is directly associated with the ith variable, and that portion of variance shared among the independent variables which is to be associated with the ith variable, by using the simpler procedure as given by formula [2], with the addition of a further, simple computation. First, the $\beta_{oi}r_{oi}$ product is found, yielding the total direct and shared variance associated with the ith variable. Then $\beta_{oi}$ is squared, producing the variance directly associated with the ith variable. This figure is now subtracted from the $\beta_{oi}r_{oi}$ product, the difference being that portion of the variance, shared by the independent variables, which is associated with the ith variable.

It is believed that this method provides two advantages over the procedures represented by either formula [1] or [2] alone: first, it yields maximum information concerning a given variable's participation in the variance accounted for in the criterion; and second, it materially reduces the number of computations necessary to secure that amount of information.

## REFERENCES

1. Engelhart, Max D. "The Technique of Path Coefficients," Psychometrika, I (1936), 287-93.
2. Guilford, J. P. Fundamentals of Statistics, 3rd Edition (New York: McGraw-Hill Book Co., 1956).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 3, March 1960)

# ON A TEST OF WHETHER TWO SETS OF OBSERVATIONS BELONG TO A CERTAIN CLASS OF DISTRIBUTIONS WHEN THE RANDOM VARIABLES ARE NOT INDEPENDENT

H. P. KUANG*
University of Minnesota

THERE EXIST several techniques which have been developed in recent years to test the hypothesis that two samples have been drawn from the same population. These developments were based on signs, ranks, or the randomization principle. For instance, to test the hypothesis that the two samples are from the same population, Mood considered a test based on signs while Dixon presented a statistic based on ranks. Utilizing the randomization principle, Pitman proposed the difference of sample means as a criterion, large values being significant. Wald and Wolfowitz devised the total number of runs in a sequence for testing the same hypothesis that two samples (not necessarily of equal size) have been drawn from a common population. The tests developed are essentially limited to the continuous and independent variables. We shall, however, treat the problem without these restrictions. It is the purpose of this paper to present a probabilistic inequality which may be used to determine whether two variables are contained in a certain set of distributions without the assumptions of continuity and independence.

Let us consider the two random variables X and Y which are not independent. Suppose that each of the variables is divided into a finite number of parts $A_1$-----$A_r$, and let the corresponding discrete marginal distribtuions of X and Y be $F_1 = \{p_i\}$ and $F_2 = \{q_i\}$ $(i = 1 \text{ -- } r)$ respectively, so that $p_i = P(X = A_i)$, $q_i = P(Y = A_i)$ and

$$\sum_{i=1}^{r} p_i = 1, \quad \sum_{i=1}^{r} q_i = 1 \quad .$$

Let, further, the corresponding observations X and Y be $\{n_i\}$ and $\{m_i\}$ $(i = 1 \text{ --- } r)$ respectively, so that the sample values $n_i$ and $m_i$ belong to the set $A_i$ and

$$\sum_{i=1}^{r} n_i = n, \quad \sum_{i=1}^{r} m_i = m.$$

Any set of $A_i$ carries the mass $\frac{n_i}{n}$ in the distribution of X and the mass $\frac{m_i}{m}$ in the distribution of Y. The object here is to decide on the basis of two sets of observations $\{n_i\}$ and $\{m_i\}$ $(i = 1 \text{ --- } r)$ whether the two distributions $\mathbf{F}_1$ and $\mathbf{F}_2$ belong to the same class of distributions.

Using the Minkowski inequality, the author has proved in his unpublished dissertation (University of Minnesota) that if the two distributions $\mathbf{F}_1$ and $\mathbf{F}_2$ belong to a certain set of distributions, the following probabilistic inequality must hold:

$$P\left\{\phi(F_1, F_2) \geqslant 1 - \frac{\epsilon^2}{2}\right\} > 1 - \frac{r^2 + r - 1}{\epsilon^4} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)^4$$

where $\phi(F_1 F_2) = \sum_{i=1}^{r} \sqrt{\frac{n_i}{n} \frac{m_i}{m}}$ and $\epsilon$ is any positive numerical value. It should be noted that the derivation of the inequality was based on the conditions that $n > r$, $m > r$, and that when $n_i \leqslant 1$ or when $m_i \leqslant 1$, $i \neq 1$. By use of the above result we can then evaluate an upper bound of the probability in terms of a pre-assigned risk, $\alpha$. It is easily seen that if $\mathbf{F}_1$ and $\mathbf{F}_2$ have the same distribution, we get

$$\phi(F_1 F_2) \geqslant 1 - \tfrac{1}{2} \sqrt{\frac{r^2 + r - 1}{\alpha}} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)^2 \quad .$$

As a numerical illustration, consider the two distributions of grades which were received by students enrolled in courses in calculus and differential equations at a university. The number of students assigned to each of five grades A, B, C, D, and F for these two courses were as follows:

---

*The author is currently Professor of Mathematics at The North Carolina Agricultural and Technical College in Greensboro, North Carolina.

| | A | B | C | D | F | Total |
|---|---|---|---|---|---|---|
| Calculus | 12 | 31 | 52 | 57 | 48 | 200 |
| Differential Equations | 2 | 20 | 44 | 50 | 44 | 160 |

The calculus was a prerequisite for the course in differential equations. These two courses were composed of not only the same individual students but also some twins, brothers and sisters. Very few of these students were, however, transfers. It frequently happened in mathematics courses that students with a good record in calculus usually made a good grade in differential equations. On the other hand, students with a poor grade in calculus were ordinarily unable to do satisfactory work in differential equations. For these reasons, the two distributions of grades are not independent. Accordingly, it is difficult to see how the ordinary test based on the assumption of independence could possibly be used. We now wish to know whether the two distributions of grades have come from the same population without the customary restriction of independence. The table at the bottom of this page gives the numerical calculations of $\Phi(F_1 F_2)$.

From the last row of the table, we obtain

$$\Phi(F_1 F_2) = \sum_{i=1}^{5} \sqrt{\frac{n_i}{n} \ \frac{m_i}{m}} = .989 \quad .$$

If we use a risk, $\alpha$, of .05, say, we get

$$\epsilon^2 = \sqrt{\frac{25+5-1}{.05}} \left( \frac{1}{\sqrt{200}} \quad \frac{1}{\sqrt{160}} \right)^2 = .5395 \ .$$

Thus $\Phi(F_1 F_2) \geqslant 1 - \frac{\epsilon^2}{2} = .730$ .

Therefore, the two distributions of grades in question belong to the same set of distributions.

Finally, it may be added that if the random variables X and Y are assumed to be independent, we then have

$$P\left\{ \Phi(F_1, F_2) \geqslant 1 - \frac{\epsilon^2}{2} \right\} \geqslant 1 - \frac{16(r^2 + r - 1)}{\epsilon^4} \left( \frac{1}{n^2} + \frac{1}{m^2} \right) + \frac{256(r^2 + r - 1)^2}{\epsilon^8 \ n^2 \ m^2}$$

when $F_1 = F_2$ .

It can be readily shown that the previous result (without the assumption of independence) is a better bound of probability when the following inequality holds:

$$15(n^2 + m^2) - 4\sqrt{nm} \, (\sqrt{n} + \sqrt{m})^2 > \frac{256(r^2+r-1)}{\epsilon^4}$$

| | A | B | C | D | F |
|---|---|---|---|---|---|
| Calculus $\frac{n_i}{n}$ | .060 | .155 | .260 | .285 | .240 |
| Differential Equation $\frac{m_i}{m}$ | .013 | .125 | .275 | .312 | .275 |
| $\sqrt{\frac{n_i}{n} \ \frac{m_i}{m}}$ | .028 | .139 | .267 | .298 | .257 |

## REFERENCES

1. Dixon, W. J. "A Criterion for Testing the Hypothesis that Two Samples are from the Same Population," Annals of Mathematical Statistics, XI (1940), pp. 199-204.
2. Matusita, K. "Decision Rules Based on the Distance for Problems of Fit, Two Samples and Estimation," Annals of Mathematical Statistics, XXVI (1955), pp. 631-40.
3. Mood, A. M. Introduction to the Theory of Statistics (New York: McGraw-Hill, 1950), pp. 390-95.
4. Pitman, J. G. "Significance Tests Which May be Applied to Samples from Any Population," Supplement Royal Statistical Society Journal, IV (1937), pp. 119-30.
5. Wald, A. and Wolfowitz, J. "On a Test Whether Two Samples are from the Same Population," Annals of Mathematical Statistics, XI (1940), pp. 147-62.

# The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1960 by
DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 4, June 1960)

# THE PERCEPTION OF MUSIC SYMBOLS IN MUSIC READING BY NORMAL CHILDREN AND BY CHILDREN GIFTED MUSICALLY*

ROBERT G. PETZOLD
University of Wisconsin

## I. Introduction

### Introductory Statement

THE COMPLEX patterns of sounds which are called music may, subject to certain limitations, be indicated with a relatively high degree of accuracy through the use of an equally complex system of musical notation. This system utilizes three types of symbols: (1) those which indicate the pitch of musical sounds; (2) those which indicate the temporal duration of musical sounds; and (3) those which indicate certain related aspects of musical interpretation such as tempo, dynamics, phrasing, and style of performance. Music reading is a highly complex process because it involves not only the recognition and interpretation of these various symbols, but frequently requires a simultaneous response in terms of vocal or instrumental performance.

A study of the professional literature shows that several terms may be used to identify the process of reading music, or to identify a specific kind of level of music-reading activity. The present study defines music reading as the process of reading and interpreting the various kinds of music symbols and converting these symbols into sound. This does not imply accurate reproduction at first sight but does suggest that the individual needs to acquire sufficient skill in understanding and interpreting the meaning of the symbols so as to ultimately reach a high level of correct response. The symbols become a guide for action, with the individual making frequent value judgments regarding the accuracy of his performance, and consequently changing that performance to agree with those judgments.

Music educators are agreed as to the significance of music reading skills in the musical development of the individual. This type of skill is considered an essential element of musical understanding and appreciation, as well as of independent musical performance. Consequently, the music program of the school gives considerable emphasis to a variety of activities designed to facilitate the development of this skill. However, there has been little systematic research concerned with the several aspects of music reading although considerable discussion has taken place regarding the following:

1. Adequate identification of the learning process as it applies to music reading.
2. Development of effective procedures for teaching music reading.
3. Designing effective instruments to measure and evaluate music reading competence.
4. Determining reasonable levels of music reading competence for children representing various grade levels and various degrees of musical ability.

### Purpose of the Study

The present study seeks to develop experimental learning situations which would provide significant information regarding the perception of music symbols by children of average musical ability and children gifted musically at the fourth and sixth grade levels. The process of reading music appears to depend upon three perceptual levels: 1) the auditory perception of musical sounds; 2) the visual perception of musical symbols; and 3) the integrative, internalized process through which the individual organizes previous auditory and visual perceptions of given stimuli in order to react to these same, or similar, stimuli as they appear in new learning situations. Each of these three perceptual levels will be explored experimentally, comparisons will be made in terms of the two groups of children at each of the two grade levels, and hypotheses will be tested.

---

*The research reported herein was performed pursuant to a contract with the U. S. Office of Education, Department of Health, Education, and Welfare. The writer wishes to acknowledge the assistance and cooperation given by the children of the Madison, Wisconsin, elementary schools; their special music teachers; and Dr. Philip Falk, Superintendent of Schools, and Mr. LeRoy Klose, Supervisor of Music.

This study deals primarily with reading those music symbols which indicate pitch; other kinds of symbols are to be considered in future investigations. Many music educators recommend that music reading skills can be developed by learning a "vocabulary" of tonal patterns, or configurations, which can then be applied to new situations. The basic assumption underlying this approach is that such tonal configurations, when learned by the child, retain their visual identity in different contexts. While the professional literature shows almost complete agreement that music reading is a response to a group of notes rather than to single notes, there is less agreement as to the procedures to be followed. Is there a standard list of tonal configurations which should be learned apart from the song materials themselves, or should the tonal configurations to be learned be taken directly out of the song materials for study and then returned to their original context?

The hypotheses of the study are as follows:

1. Normal children perceive and act upon music symbols in essentially the same way as do children gifted musically.
2. Certain tonal configurations which are readily perceived in isolation by both groups of children tend to retain their identity when presented in the more usual context of a song.
3. There are no significant differences between the ability to perceive a tonal configuration aurally and the visual perception of these same configurations.
4. Age, previous musical experience and musical ability are significant factors in developing skill in perceiving and reacting to music symbols.

Procedures of the Study

The procedures followed in the study are briefly outlined below and will be discussed in detail in a later section of this report.

1. Analysis of song materials to identify common tonal configurations which may be used as test items.
2. Selection of subjects for the study and collection of certain personal data on each subject.
3. Construction of an initial test and administration of this test to approximately 100 subjects selected at random from the fourth fifth and sixth grade pupils. This constitutes Phase I of the study and will provide one means of examining perception as the visual and aural identification and recognition of tonal configurations. It will also establish, in part, the basis for selecting learning tasks appropriate to Phase II of the study and provide an opportunity to study the techniques appropriate for the collection and analysis of the data.
4. Construction of the test materials for Phase II of the study and administration of the test to 150 average and musically gifted children in both fourth and sixth grades. Phase II is primarily concerned with perception as an integrative, internalized process through which the individual relates his auditory and visual perceptions of tonal configurations to his perception of these same configurations as they appear in a different setting.
5. Final analysis of the data, with small group comparisons between grades and between groups being made.

The testing carried on in the study is all individual, with the data gathered by tape-recording the pupil responses in the testing situations and later subjecting these recordings to detailed analysis. Tape recording the pupil responses in the testing situation eliminated the need for immediate scoring and tabulating and insured accuracy of the data since questionable responses could readily be repeated for verification.

Throughout the present study the level of skill in music reading is measured in terms of a vocal response to the stimuli. When an instrumental response is used, it is difficult to judge whether the subject really understands and interprets the musical symbols being read, or whether the response is mechanical and is being made in terms of fingering patterns. Vocal response was considered to be the most practical for the study because all children had had extensive experience in singing as a part of the regular music program. The use of an instrumental response would have further restricted the study to a relatively small selective group and thus defeated the primary purpose of the investigation.

## II. Summary of Related Information

The writer made an extensive analysis of the professional literature related to the basic problem of music reading. The problem of developing skill in music reading has been a fertile topic for discussion in the professional literature for a period of many years. Music educators have offered many suggestions and recommendations with reference to appropriate procedures for the development of this highly specialized skill. The objectives of the music program in the schools, as stated by authorities in the field, have always mentioned the need for acquiring a certain level of music reading skill. There exists, however, a considerable amount of uncertainty regarding exactly why children need to learn how to read music, as well as a general tendency to avoid making specific recommendations regarding effective and adequate procedures which might be followed. One valid reason for the vagueness with which these two points are discussed may well be due to the lack of adequate research information.

The literature concerned with the topic of music reading may be grouped into three broad categories:

1. The historical background of musical notation and of syllabic devices for interpreting musical notation.
2. Professional literature which includes courses of study, specialized music education methods books, and the many articles in professional journals which report the opinions, experiences, and recommendations based upon personal experience.
3. Research literature.

For the sake of brevity, the writer will present only a partial summary of certain pertinent research literature.

Analysis of the research in the specific area of music reading reveals the following pertinent information. One group of studies has approached the problem through a detailed analysis of the eye movements of persons engaged in reading music. Jacobson (4) following the procedures of Gray and Buswell in the field of language reading, carried forward a detailed photographic analysis of eye movements as a means of identifying the various stages in the development of music reading ability. The section of the report which is pertinent to the present study deals with the reading of vocal music by three college students and three sixth-grade pupils. A detailed analysis of the performance of each subject is presented and the results are briefly summarized here. The less mature reader required pauses almost twice as great as the mature reader but the number of pauses did not differ significantly. The span of recognition, which included both words and notes, seldom exceeded two notes for the mature reader and was smaller than the span of the earlier instrumental reading. The smaller intervals were read correctly more often than longer intervals, use of tonic sol-fa syllables resulted in longer pauses than when words were used; and sol-fa syllables were found to be of no benefit to the subjects. The author offers several suggested applications of the results of this study to classroom instruction which seem to indicate that new and simplified reading materials are needed to develop the immature reader.

Ortmann (7) summarized one of his investigations by pointing out that the span of vision depends a great deal upon the placement and pattern of the notes. He found the field of clear vision to be a circle varying between 1/4 and 1/3 inch in diameter with no difficulty encountered in reacting to notes within this area. The chief determinants of difficulty in note reading, not involving actual eye movements, are: a) the number of notes in the field of clear vision, b) the area covered by the distribution of the notes, c) the number of linear dimensions involved (i. e., combinations of vertical horizontal, or oblique), d) the complexity or symmetry of the note pattern, and e) the meaning of the note group as a melodic and/or harmonic pattern. He points out that difficulties in note reading, for the normal pupil, result from an inability to group a distribution of notes into larger perceptual units. Familiarity with musical notation is quite obviously a significant factor in accuracy of reading.

A third study of eye movements in music reading was carried on by Weaver (10) using trained musicians, with the piano as the performance medium. His experimental reading materials were harmonic, melodic, and single melody with supporting chords. The rhythmic element was retained in all types as a means of accurately measuring the number and duration of reading pauses. The results indicate that note reading relates more closely to reading algebraic formulae than to language reading, assuming that one musical note is equivalent to one word. The average number of notes played per reading pause varied between one and two, with spacing of notes and harmonic grouping being important factors. The duration of reading pauses depended upon the type of musical material being read, with those for harmonic and melodic material reliably shorter than those required in reading an accompanied melody. A higher correlation exists between reading time and pause duration than between reading time and the number of pauses. Weaver also found that performance in his study provided little basis for distinguishing types of readers because of the individual differences in performing the different selections. It is significant that he identified the average perceptual span as varying between three and five notes. A companion study by Weaver and Van Nuys (11) was particularly concerned with the relationships between the rhythmic and melodic elements of music as they affected occular and manual behavior. The immediate memory span was used as a criterion of difficulty in determining the eye-hand span, with immediate memory span defined as the amount of material which could be visually explored, perceptually organized and retained until played. The results showed that memory span decreases as the complexity of note relations increases for both rhythm and melody. The memory span for a combination of a simple rhythm and melody is only slightly smaller than the span for each alone. In more complex combinations, the melodic element seems to prove more difficult than the rhythmic. Melodic factors limit memory span whenever the melody is not extremely simple; rhythmic factors limit the rate of reading when complex rhythms are used. If music reading skill is to be developed so more rapid responses are made, individuals need to improve in the ability to apprehend pitch patterns as stable melodic segments and to grasp rhythms as rhythmic figures.

Bean (1) sought to determine the complexity of the musical pattern which could be perceived at one fixation of the eyes by individuals with varying amounts of training. He also studied the effects of tachistoscopic training on the span of perception. The investigation showed that the ability to grasp a musical pattern of three or more notes at a glance was essential to rapid, accurate reading. The perceptual unit is determined by the duration of notes, the distribution of notes on the staff, the closeness of note groupings, and the irregularity or unusual characteristics of patterns. These factors all operate to determine the ease with which material is grasped by the reader. Disjunct melodic patterns, rhythmic patterns which could not readily be subdivided into logical sub-groups, unfamiliar musical material, and uneven distribution of chord tones appear more difficult to perceive than other types of material.

Christ (2) employed tachistoscopic training as a means of improving the ability to perceive and reproduce rhythmic units. The results showed that such drill yielded significant gains but no attempt was made to determine the permanence of such drill or to relate such drill to rhythmic reading of typical music. Stokes (9) investigated the effect of tachistoscopic training on reading melodic material. The results, however, showed that the gains on a post-test could not be considered due to the influence of the specific training program alone. An interesting instrument for testing the sight-reading performance of instrumental performers was constructed by Stokes and contains several valuable suggestions. The testing results show, however, that tachistoscopic training had no effect upon general achievement in reading music and that there was little evidence of carry-over.

Smiley (8) investigated the music reading "vocabulary" in basic music texts at the fourth-grade level and found the total number of tonal configurations to be far too many for the child to respond to readily. The 351 configurations could, on the basis of frequency of use, be reduced to approximately 20, each containing from two to five notes. Further analysis led Smiley to conclude that it might be more valuable for the child to learn a basic list of two-note configurations of seconds and thirds. The procedures suggested by Smiley for identifying tonal configurations were a valuable guide for the analysis made by the writer.

This reported research in music reading has been primarily concerned with certain aspects of visual perception or of physical behavior during the process of reading music. Such research has not yet provided an adequate basis for prescribing instructional procedures to develop musical concepts which may be readily utilized in a total musical setting. Apart from a few reports of actual classroom experimentation in the problem area, there is little information regarding the ways in which children of elementary school age react to musical notation because the majority of the studies have used older subjects. The survey of literature again re-emphasizes the need for continued, systematic investigation of this, and other problems in music education.

## III. Procedures of the Study

### A. Analysis of Songs

The initial step taken in the study was a detailed analysis of 326 songs selected at random from eighteen of the basic song-texts used in grades three through six. Since the subjects for the study were to be selected from the fourth, fifth, and sixth grades, it was necessary to use test items with which they would be reasonably familiar as well as items which would be discriminating. The purpose of the analysis was to identify tonal configurations and select test items on the basis of the frequency of appearance of certain tonal configurations in songs.

The procedures followed in identifying the tonal configurations were suggested, in part, by Smiley (8). Other suggestions were obtained from the teachers' manuals to the basic song-texts and from professional books and articles whose authors advocated the "pattern" approach to music reading. The following rules were set up to serve as a guide in defining a tonal configuration:

1. Thirty page numbers, chosen from a table of random numbers, were set up for each of the eighteen song-texts. Although only eighteen songs in each book were to be analyzed, the additional songs were necessary as alternates because of the decision to eliminate songs in a minor mode or songs containing frequent chromatic or modulatory passages.
2. All types of songs were included in the analysis, no distinction being made between specific reading songs, rote songs, or rote-note songs.
3. Single intervals were generally not considered as tonal configurations since they do not imply any melodic flow or movement.
4. The interval between musical phrases was not usually included.
5. The interval within a phrase which appeared between the end of one distinct configuration and the beginning of another configuration was not in-

cluded unless the text showed that such an interval occurred in the middle of a word.

6. The interval between readily identifiable sequences was not included, each sequence being treated as a separate configuration.

7. Repetitions of single notes were considered as single notes; do-do-re-re-do-do- was analyzed as do-re-do. The decision to exclude repeated tones from configurations was made in order to simplify the analysis even though such repetitions occur frequently in songs.

8. In part songs, each part was analyzed separately only in those sections of the song where the soprano and alto parts were different.

9. In songs containing repetitions, only the new ending was analyzed, not the entire repetition.

10. Frequently the rhythmic structure within the measure or phrase could be used to assist in the identification of a complete tonal configuration.

11. The text of the song itself, particularly with reference to words of more than one syllable, could frequently be depended upon to be of assistance in accurately identifying the tonal configurations.

12. Within these guides, the decision regarding the length of the configurations had to be made by the investigator in terms of what the child could be expected to grasp. As a result, tonal configurations in excess of five or six tones were rarely identified as such.

Table I summarizes the numerical results of the analysis. The total number of different tonal configurations identified in the 326 songs was 558. The total frequency of these 558 configurations was 4575, with a frequency range of 1 to 213. It is significant that 47 tonal configurations, or only 8.42 percent of the total number identified, with a frequency range of 20 to 213, had a combined frequency count of 2585, or 56.50 percent of the total frequency count. Even more significant is the fact that 22 configurations, or not quite four percent of the total, with a frequency range of 44 to 213, had a combined frequency of 1939, or 42.38 percent of the total frequency count. Tonal configurations with a frequency range of 1 to 9, accounted for 460, or 82.44 percent of the total number of configurations, but had a combined frequency count of only 1327, or 29 percent of the total frequency count. The data points up the fact that there exists a relatively small number of tonal configurations which appear repeatedly in the songs children sing.

Had songs in a minor mode been analyzed, as well as songs containing chromatic and modulatory passages, the total number of tonal configurations would have been considerably in excess of 558. The situation is still further complicated by the fact that each one of these 558 configurations can be written in at least nine different major keys commonly used in these song-texts. Since each configuration has, therefore, nine different notational systems within which it could appear, the total possible number of 5022 configurations appears to present a very formidable problem to the music educator faced with the task of identifying effective methods for developing skill in music reading. Fortunately, the general outline or shape of a configuration remains constant from one tonality to another despite differences in actual notation. However, this problem of perceiving and identifying configurations when they are presented in several different tonalities needs careful investigation since it is frequently dismissed as being relatively unimportant.

Since the selection of test items was to be based, in part, upon the frequency of appearance in songs, the twenty-two most frequently used configurations are listed in Table II. For purposes of comparison, the twenty-two most frequently used configurations identified in the analysis made by Smiley are also presented. These twenty-two configurations represented 3.94 percent of the total number of configurations identified by the writer and 6.1 percent of the total number of configurations identified by Smiley. It is significant to note that although Smiley confined his analysis to fourth-grade songs identified as reading songs, his findings closely agree with those of the present investigator. Nineteen configurations are common to both lists, with each list containing only three configurations not found in the companion list. These configurations are identified by an asterisk. It would be reasonable to expect that analyses of this type, carried on by two investigators operating independently but following similar criteria of identification, would show comparable results.

In order to insure an adequate number of tonal configurations from which test items could be constructed, an additional 25 configurations were listed. These configurations, with a frequency range of 20 to 42, accounted for 646, or 14.12 percent of the total frequency count and are listed in Table III. It is significant to note that there again exists considerable agreement between the analysis made by Smiley and that made by the writer. Sixteen, or 64% of these 25 configurations appear on both lists. The 47 tonal configurations, with a frequency range of 20 to 213, which have been identified by the writer, included 35 configurations having a similar frequency range as identified by Smiley. This represents a concurrence of identification for 75 percent of the highest ranking items. No attempt was made to compare results of this analysis with the tonal configurations identified in professional books and teacher's guides to the basic song-texts since it was not considered significant to this particular study.

These 47 configurations were considered a suf-

TABLE I

THE FREQUENCY COUNT OF 558 TONAL CONFIGURATIONS

| Frequency Range | Number of Configurations | Percent of Total Count | Combined Frequency | Percent of Total Frequency |
|---|---|---|---|---|
| 44-213 | 22 | 3.94 | 1939 | 42.38 |
| 20- 43 | 25 | 4.48 | 646 | 14.12 |
| 10- 19 | 51 | 9.14 | 663 | 14.49 |
| 1- 9 | 460 | 82.44 | 1327 | 29.01 |
| Totals | 558 | 100.00 | 4575 | 100.00 |

## TABLE II

### THE 22 MOST FREQUENTLY USED TONAL CONFIGURATIONS AS IDENTIFIED IN THE PRESENT STUDY AND BY SMILEY

| Present Study | | | Smiley's Study | | |
|---|---|---|---|---|---|
| Configuration | Frequency | % Total Frequency | Configuration | Frequency | % Total Frequency |
| so do | 213 | .047 | so do | 191 | .041 |
| mi re do | 167 | .037 | so la so | 135 | .029 |
| so la so | 136 | .030 | do so | 102 | .022 |
| mi fa so | 115 | .025 | mi re do | 99 | .021 |
| do re mi | 110 | .024 | so mi | 91 | .020 |
| do ti la so | 106 | .023 | mi so | 88 | .019 |
| do mi so | 93 | .020 | do ti la so | 80 | .017 |
| so mi do | 93 | .020 | do ti do | 73 | .016 |
| so fa mi | 92 | .020 | do re do* | 72 | .016 |
| so mi | 90 | .020 | so mi do | 70 | .015 |
| mi so | 76 | .017 | do mi so | 69 | .015 |
| so la ti do | 76 | .017 | fa mi re do | 69 | .015 |
| do la so | 69 | .015 | do re mi | 66 | .014 |
| do so | 68 | .015 | mi do | 65 | .014 |
| do ti do | 65 | .014 | so fa mi | 64 | .014 |
| so fa mi re do | 60 | .013 | do la so | 60 | .013 |
| so mi so* | 58 | .013 | so la ti do | 59 | .013 |
| mi do | 55 | .012 | mi fa so | 58 | .013 |
| do mi | 55 | .012 | so fa mi re do | 58 | .013 |
| fa mi re do | 54 | .012 | so la* | 53 | .011 |
| so do mi* | 44 | .010 | do la* | 46 | .010 |
| re mi fa* | 44 | .010 | do mi | 43 | .009 |
| Total | 1939 | .424 | Total | 1711 | .369 |

* Configurations not found in companion list.

**The figures in this column are to be interpreted as follows: .047 is 4.7 percent, etc.

## TABLE III

## FREQUENCY ARRANGEMENT OF 25 TONAL CONFIGURATIONS HAVING A FREQUENCY COUNT OF 20 TO 39

| Configuration | Frequency | Configuration | Frequency |
|---|---|---|---|
| do re mi fa so* | 39 | so mi* | 25 |
| do re do | 36 | so do* | 24 |
| do re mi do* | 31 | so fa re do | 23 |
| mi fa mi* | 30 | fa re* | 23 |
| fa mi re* | 30 | la ti do* | 23 |
| do do | 29 | do la do | 22 |
| do ti la* | 28 | mi so do | 22 |
| re mi re do* | 27 | so la so fa mi | 22 |
| do mi do* | 27 | la so fa mi* | 21 |
| mi re mi* | 27 | so do so | 20 |
| do re mi re do* | 26 | re fa mi re do | 20 |
| do so do | 26 | so fa mi re* | 20 |
| mi so mi | 25 | Total | 616 |

*These items were also identified by Smiley as having a frequency count within this range of 20 to 39.

## TABLE IV

## DISTRIBUTION OF THE SUBJECTS PARTICIPATING IN PHASE I

| School | 4th Grade | | 5th Grade | | 6th Grade | | |
|---|---|---|---|---|---|---|---|
| | Boys | Girls | Boys | Girls | Boys | Girls | Totals |
| School A | 5 | 5 | 5 | 5 | 5 | 6 | 31 |
| School B | 4 | 5 | 4 | 5 | 5 | 5 | 28 |
| School C | 5 | 6 | 5 | 4 | 5 | 5 | 30 |
| Total | 14 | 16 | 14 | 14 | 15 | 16 | 89 |

ficient number from which to select test items and any further analysis of the data was unnecessary at this time. The analysis was of considerable value in emphasizing a basic problem which the writer feels needs clarification if music educators are to continue to support the "pattern" approach to music reading. It involves establishing objective, rather than subjective criteria for accurately identifying tonal configurations. Such objective criteria might well be established if the procedures suggested in the research concerned with the photographing of eye-movements were to be applied to the problem. Certainly such an investigation would also need to consider the ways in which the perception of any given tonal configuration might be significantly influenced by the factors of rhythmic structure, the accompanying text of a song line, and the overall movement and direction of the melodic line.

## B. Selection of the Sample

The subjects for this study were fourth-, fifth- and sixth-grade pupils selected from three of the twenty-two public elementary schools in Madison, Wisconsin. Limitations of time made it impossible to derive a sample which would be representative of the total population of these three grades, and it was, therefore, decided to select the subjects from three schools, these schools being representative of the overall school system. The criteria for selection were school size, geographic location and socio-economic background of the children. School A is a large school, with an enrollment of 841, located in the east area of the city where pupils are drawn largely from the semi-professional group. School B is a small school, with an enrollment of 478, located in the central area of the city where pupils came from the skilled and semi-skilled group. School C is of medium size, enrolling 660 pupils, and located in the west section of the city in an area largely composed of the professional group.

1. Sample for Phase I—The study is divided into two distinct parts, each utilizing a different sample. Phase I utilized a "normal" population of pupils selected at random from the fourth, fifth and sixth grades of each of three schools and undifferentiated in terms of musical ability. Class lists were secured and an equal number of pupils from each grade in each school were selected at random, with equal numbers of boys and girls. Although more than 100 children were tested in Phase I, usable data was secured for 89 pupils. Table IV shows the distribution by school and by grade.

The data for several subjects was judged unusable because they were unable to make the type of response called for by the test, a singing response to aural and visual stimuli. It was considered unwise to administer the entire test to children for whom there could be little success on any item.

2. Sample for Phase III—Phase II of the study utilized two reasonably well-defined segments of this "normal" population of the three selected schools, pupils identified as those gifted musically and those possessing average musical ability. Subjects were selected from grades four and six for each of these two groups. The initial step taken in the identification of these two groups was to establish criteria which could be followed. There is considerable discussion in professional literature regarding the gifted child, but there exist few criteria for identification in the field of music. Usually such judgments are made on the basis of one or more of the following: a) excellence in musical performance; b) demonstrated knowledge of musical information and concepts as measured by achievement tests; c) results of music aptitude or talent tests; d) grades in music subjects; and e) rating scales constructed for the purpose.

The schools participating in this study do not follow a standard testing program to establish levels of musical achievement and there was no readily available data which could be used for identifying possible gifted children. A preliminary investigation of available professional literature had failed to yield many concrete suggestions regarding procedures which could be used to identify musically gifted children other than the criteria which have already been mentioned. Each school had an experienced special music teacher, and it was felt that this person would be capable of identifying gifted pupils if certain criteria were made available. Consequently, a rating sheet was designed by the writer and included twelve items considered by music educators as important factors in determining musical accomplishment of elementary school pupils. (See next page for rating scale.) The rating scale was submitted to a group of music teachers for comments and suggestions and a final form was drafted. The music specialist in each school was then asked to evaluate all fourth- and sixth-grade pupils, the rating sheets were then scored and tabulated. Obtaining evaluations for all pupils made it possible to identify for each grade in each school, three groups of pupils: a) those receiving the highest scores; b) those receiving the lowest scores; and c) those in the middle of the distribution. This procedure led to a more accurate identification of the "average" group since it could be assumed that they were to be found between the upper and lower groups. For each grade in each school the upper ten percent of the total distribution was tentatively selected as being musically gifted. The mean rating score and standard deviation was calculated for each grade in each school and the "average" group was tentatively defined as being within the limits of one-half a standard deviation on

## PUPIL RATING SCALE

________________________ (Name) ________________________ (School) ____________ (Grade)

(Instructions: Please check the response which best describes the pupil's ability or attitude in terms of his performance on each item.)

### A. GENERAL SINGING ABILITY

1. The pupil is able to match individual tones and intervals:
   ______Always
   ______Usually
   ______Occasionally
   ______Seldom
   ______Never

2. The pupil is able to sing a phrase of a song he knows:
   ______But never succeeds, even with help from the piano or teacher
   ______Seldom succeeds, even with help
   ______Needs frequent help
   ______With only occasional help
   ______Independently, with no help

3. In the independent (solo) singing of a song learned by the class, the pupil:
   ______Sings the complete song without help from the teacher or piano
   ______Sings the complete song but needs occasional help
   ______Sings the complete song but needs frequent help
   ______Is seldom able to sing the complete song, even with help
   ______Is never able to sing the complete song, even with help

4. In the independent (solo) singing of a song learned by the class, the pupil:
   ______Almost never sings in tune, usually fails to adjust the pitch when helped
   ______Seldom sings in tune, occasionally is able to adjust the pitch with help
   ______Frequently sings out-of-tune, usually adjusts when given help
   ______Occasionally sings out-of-tune but always adjusts when helped
   ______Almost always maintains correct intonation throughout without help

5. In the independent singing of a song learned by the class, the pupil:
   ______Never makes interval errors, never loses the tonality or key-sense
   ______May make occasional interval errors but regains tonality without help
   ______Makes interval errors and needs help to return to proper tonality
   ______Frequently loses the tonality and does not always respond to help
   ______Almost always loses the tonality and usually fails to respond to help

6. The pupil's singing, either alone or with the class, is characterized by an enthusiastic, eager, whole-hearted and controlled response:
   ______Almost never (0-20% of the time)
   ______Occasionally (21-40% of the time)
   ______Frequently (41-60% of the time)
   ______Usually (61-80% of the time)
   ______Always (81-100% of the time)

### B. RHYTHMIC RESPONSIVENESS

7. The pupil, when asked to identify the underlying beat or meter of the song by hearing it, does so immediately and accurately:
   ______Always
   ______Usually
   ______Occasionally
   ______Seldom
   ______Never

8. The pupil, when asked to repeat common rhythmic patterns which he hears, does immediately and accurately:

______ Never
______ Seldom
______ Occasionally
______ Usually
______ Always

## C. MUSICAL KNOWLEDGES AND SKILLS

9. When participating in appropriate music reading activities, the pupil's response shows that he is:

______ A leader, capable of much independent music reading
______ Partially independent, voice always moves in direction of notes and is frequently accurate
______ Occasionally able to control the direction of the voice by using notes
______ Seldom is able to react to the notation, learns by rote most of the time
______ Seems to be unaware of the meaning of the notation, cannot differentiate one symbol from another

10. The pupil demonstrates a knowledge of common musical symbols and terms used in his grade that is:

______ Poor
______ Fair
______ Average
______ Good
______ Superior

## D. GENERAL INTEREST IN MUSICAL ACTIVITIES

11. In terms of the singing activities carried on, the pupil is:

______ Very interested and responsive
______ Usually interested and fairly responsive
______ Passive, does not put forth much effort but does not resist
______ Slightly negative, frequently does not participate at all
______ Has a strong negative reaction, usually does not participate

12. In terms of the listening and creative activities carried on, the pupil is:

______ Strongly negative, fails to pay attention or show interest
______ Slightly negative, frequently does not show interest
______ Passive, shows neither a negative or positive reaction
______ Is usually very interested and fairly responsive
______ Very interested and responsive

either side of the mean. The ratings for each school were kept separate since it obviously could not be assumed that all three teachers would interpret the criteria in the same way.

In addition to the teacher ratings, a second criterion measure was used to identify the musically gifted. The Kwalwasser Music Talent Test, Form B, was administered to all fourth- and sixth-grade pupils in the three schools. The decision to administer this particular test was based on the practical necessity of having a standardized talent test which could be given within a single twenty-five minute period, the length of the music period in the schools. The test was administered to 438 pupils and the data is summarized in Table V.

The mean score for each sixth grade is slightly higher than the mean score for each fourth grade, but it was not considered necessary to test for the significance of the difference at this time. For the combined grades in the three schools, with a total N of 438, the range of scores was 13 to 40 and the mean was 28.21.

A comparison of the data obtained from the teacher ratings and the Kwalwasser test showed that a relatively low correlation existed between the two criterion measures. Combining all pupils in both grades in each school and calculating the coefficient of correlation for each school showed the following: a) School A, r = .34; b) School B, r = .39; c) School C, r = .34. In the absence of

## TABLE V

### SUMMARY OF THE SCORES EARNED ON THE KWALWASSER TEST BY 438 4TH AND 6TH GRADE PUPILS IN THREE SCHOOLS*

| | 4th Grade | | | 6th Grade | | |
|---|---|---|---|---|---|---|
| School | Number | Score Range | Mean | Number | Score Range | Mean |
| A | 94 | 19-36 | 26.88 | 108 | 13-37 | 28.73 |
| B | 70 | 19-35 | 27.50 | 50 | 21-40 | 28.58 |
| C | 54 | 20-36 | 28.37 | 62 | 17-37 | 29.66 |
| Total | 218 | 19-36 | 27.45 | 220 | 13-40 | 28.96 |

*The highest possible score is 40.

## TABLE VI

### REPRESENTATIVE SAMPLE OF PUPILS INITIALLY SELECTED FOR THE TESTING PROGRAM OF PHASE II

| School | Total Subjects | Percent of Sample | 4th Grade | | 6th Grade | |
|---|---|---|---|---|---|---|
| | | | Average | Gifted | Average | Gifted |
| A | 66 | 44 | 20 | 10 | 21 | 15 |
| B | 38 | 25 | 15 | 7 | 8 | 8 |
| C | 47 | 31 | 14 | 10 | 13 | 10 |
| Total | 151 | 100 | 49 | 27 | 42 | 33 |

other information, one may conclude either that teacher ratings provide a more valid way of determining musical giftedness than do talent tests, or that the reverse is a more appropriate conclusion. The writer decided to consider this aspect of the problem in more detail after the data had been collected for Phases I and II. Judgments regarding those pupils who were to be classified as musically gifted were made on the basis of arbitrarily assigning equal weight to both the teacher ratings and the Kwalwasser test scores and then selecting pupils who ranked highest on both measures.

The sample for Phase II was selected on the basis of proportionate representation in terms of each school and each grade. The data showed a total of 475 fourth- and sixth-grade pupils in the three schools with School A accounting for 44 percent of the total; School B for 25 percent of the total; and School C for 31 percent of the total. These proportions were maintained in the sample and, for each school, the proportionate number of pupils enrolled in the fourth and sixth grade was also maintained. An arbitrary total of 151 subjects was established as the maximum number that could be tested within the time limits of the study. The number of gifted pupils identified by the two criterion measures was then subtracted from the total number allocated to each grade in each school and the remainder constituted the number of average pupils. These were to be selected at random from the total distribution within the limits of one-half a standard deviation on either side of the mean. Additional subjects were selected in order to insure the availability of alternates in the event a selected subject was unable to participate in the testing program. Table VI shows the number of average and gifted pupils initially selected for Phase II of the study.

### C. Testing Procedures: Phase I

Several exploratory testing sessions had provided valuable information necessary for establishing the testing procedures to be followed in the study. These will not be discussed in the report but are mentioned so the reader is aware that certain procedural decisions which were made were based upon actual test situations and were not the arbitrary decisions of the writer.

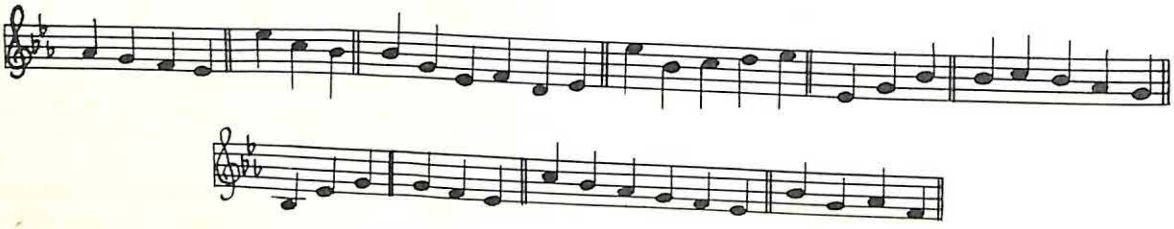
The original test contained twenty items of varying length and complexity, selected on the basis of the analysis of song materials discussed in Section IIIA. It was found that three trials of this test took too much time, and so the final form of the test was reduced to three trials of ten items. It was assumed that several trials of the same items would provide data of greater significance than would a single trial of many different items. Three-, four-, five-, and six-note items were used representing ascending and descending scale and chord patterns, either presented separately or in combinations. The construction of these test items utilized some of the suggestions of Ortmann, who was interested in defining and measuring tonal memory in music (5, 6). The items were written as quarter notes with no measure bars or time signature indicated. All items were written in the same key to reduce the number of variables in the testing situation, with the key of E♭ major considered to be most appropriate since it would keep the items within a reasonable singing range for the children. Three complete sets of 4 by 6 test cards were used with the notation on the cards duplicating, as closely as possible, the printed notation in books in order to make for easier reading. The staff contained lines spaced 1/8 inch apart which were placed in the center of the card, the length of the staff being varied according to the number of notes in the test item. The note heads were oval shaped with a lengthwise diameter of 1/8 inch and were spaced 3/8 inch apart. The treble clef sign, flats, and quarter note stems were also of suitable size. No subject reported difficulty in reading the notation, so it may be assumed that the visual stimuli were satisfactorily notated. One trial of this test is shown in Figure 1 at the top of the next page.

A second problem was concerned with the manner in which the tonality and starting pitch was to be established for each configuration. Three methods were considered: a) to sound the pitch of the starting note of each configuration; b) to sound the tonic chord of E flat major, followed by the pitch of the starting tone; or c) to sound only the key-tone E flat, or do, and let the subject locate the starting tone of the configuration from this. The literature shows that each of these methods has been used in similar testing situations, with the third method considered to be the most discriminating in terms of reading ability. This calls for knowledge of not only basic music notation but requires the subject to establish the tonal relationships internally before singing. The exploratory testing sessions established the first method as most appropriate for the study.

A third problem was concerned with the method used in presenting the items in the testing situation. The duration of the exposure time, while not a critical point in the study, needed to be uniform from individual to individual if the data were to be adequately interpreted. A memory drum with a six-second exposure time was used initially but the results showed the exposure to be too brief for an adequate reading response. Therefore, a tape recording of the entire test was made, synchronized with the individual test items which had been placed on cards. The piano was used in preference to some other medium because the children could be expected to be familiar with its sound. The timing of the tape was accomplished by setting an electric metronome so it marked every second, with the flashing light being used rather

FIGURE 1

THE 10-ITEM TEST FOR PHASE I



than the clicks to keep the tape free of unnecessary sounds.

The examiner gave the verbal cue for handing the stimulus card to the subject by saying "Card number ___" within a two-second interval. The starting pitch was sounded on the third second and sustained for two seconds. This was followed by an interval of silence during which time the subject was to sing a response to the visual presentation. A second time interval then followed which included first playing the configuration at a rate of two quarter notes per second, followed by silence during which the subject was to sing what had been played. The beginning of each item was always the same "C*ard number*___, (**sound pitch)," the "*" occurring each second. The varying time intervals are shown below:

| Length of Item | Duration for 1st Interval of Silence | Duration of 2nd Time Interval |
|---|---|---|
| 3 notes | 8 seconds | 5 seconds |
| 4 notes | 9 seconds | 6 seconds |
| 5 notes | 10 seconds | 7 seconds |
| 6 notes | 11 seconds | 8 seconds |

The subjects for Phase I of the study were selected at random from grades four, five, and six as previously described. All subjects were tested individually in a room other than a classroom. Prior to taking the test, each subject was given a brief orientation which included a statement of the purpose of the experiment together with descriptions of the function of the equipment and of the tasks to be performed. The first trial of the 10-item test included a five-item practice trial which utilized the regular test procedures and gave the subject an opportunity to ask questions about anything that was not understood. Considerable emphasis was given to the idea that hearing an item after the subject had tried to read it was designed to make it possible for the subject to recall the sound of the item the next time it was presented visually, as well as to provide a model against which he could compare the original attempt. The pupil's responses were recorded on a second tape recorder at a tape speed of 3.75 inches per second, a procedure which insured the accuracy of processing and scoring the responses.

The testing procedures for Phase I may be summarized as follows:

1. The subject was given a visual presentation of the first item, accompanied by the starting pitch of the group.
2. The subject then sang an immediate response to this visual presentation, reading the notes of the configuration.
3. This reading trial was followed by a second visual presentation of the item in combination with an aural presentation of the same item. This permitted the subject to associate the sound of the item with its visual representation as well as to provide a model against which he could compare his original attempt.
4. Immediately after this aural presentation, the subject sang what was heard, continuing to look at the notation of the item.
5. The subject then proceeded to the next item, following Steps 1, 2, 3, and 4. The testing continued in this manner until all the items had been presented.
6. The total list of items was presented a second and third time, with the items randomized for each successive presentation to eliminate the possibility of rote or serial learning.

Processing the data presented no significant problems. The tapes contained all the responses of the subjects, and these responses were transferred to individual data sheets for each subject. Correct responses for both the visual and aural stimulus were indicated by a check mark; failure to try an item or to sing only the initial notes of the item were indicated by a dash; incorrect responses, even if only a single tone was wrong, were indicated by the tonic sol-fa syllables for the entire response. Items sung out-of-tune, either in part or completely, were counted as incorrect and the exact pitch reproduction so indicated. For both the visual and aural responses, the scoring system used was the number of items for which correct responses were made in two of the three trials.

## D. Testing Procedures: Phase II

Two kinds of learning situations were constructed which differed in sequence of tasks rather than in content. These will be designated as Sequence A and Sequence B. Each sequence contained two tasks and the subject was given approximately 15 minutes to complete each task. In both sequences the two tasks were separated by a 48-hour time interval. This time interval between tasks was used for two reasons: a) the results of the testing in Phase I indicated that children of this age group were not able to give concentrated attention to tasks of this nature for sessions lasting much longer than 15 minutes; and b) the testing was to be done during the time the child met with the regular music class, these classes being scheduled on alternate days rather than daily.

These two sequences, together with their tasks may be diagrammed as follows:

| | Task I | Task II |
|---|---|---|
| Sequence A | The subject attempts to learn five tonal configurations in eight trials following the procedures established in Phase I. | The subject attempts to learn a short song constructed from the five separate items presented in Task I. |
| Sequence B | The subject attempts to learn the short song constructed from the five test items. | The subject attempts to learn the five tonal configurations which had been presented in the song. |

The material used in both sequences was the same and the testing procedures followed in each task were similar for both sequences. The major difference between the sequences was the order in which the tasks were presented to the subject.

Persons familiar with the procedures recommended in developing music reading skill will recognize that Sequence A is based on the assumption that preliminary learning of certain tonal configurations will result in more rapid learning of a song containing these configurations. Sequence B is based on the assumption that music reading skill should begin with the song and then proceed to the drill work on the musical elements contained in the song. The original design of the study included a second short song to be learned by all subjects in both sequences which would have provided much valuable data regarding the relative merits of each sequence. Unfortunately, there was not sufficient time to do the additional testing.

It was hoped that most subjects would be able to learn the test material in a single testing session. Since it could be assumed that more trials on fewer items would result in a higher percentage of success than fewer trials on more items, it was decided to use a test of five tonal configurations. Item analysis of the data on the visual responses to the items in Phase I resulted in the selection of the five items shown in Figure 2 (top of next page). The data for this item analysis will not be presented in the report since it is not considered pertinent. The techniques employed in making the item analysis were suggested by Flanagan (3).

Since the tonal configurations were again to be presented as quarter notes without time signatures or measure bars, the song to be used should not contain significantly different notations. It was decided to write a song rather than attempt to find one in song-texts which could serve the purpose. These guides were used in constructing the song:

1. It should be reasonably musical without duplicating any familiar songs.
2. Two phrases, each containing four measures, was considered an ideal length.
3. No material other than the five tonal configurations already selected was to be included. This would, however, permit the repetition of at least one tonal configuration or of single tones.
4. Notation was to use only quarter and half notes, with each tonal configuration appearing as a single measure whenever rhythmically possible.
5. The song would be in the same key as the tonal configurations and would encompass a reasonable singing range.

Ten different songs were constructed and submitted to an informal jury of composition students, theory teachers and public school music teachers for reactions and comments. There was considerable agreement that two of the ten songs were musically more interesting than the others and the test song was then selected by tossing a coin. This song, with the tonal configurations identified, is presented in Figure 3.

The subjects for Phase II of the study were selected as described earlier in the report. Each of the three schools had four groups of subjects:

FIGURE 2

THE 5-ITEM TEST FOR PHASE II

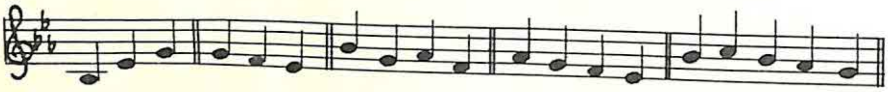
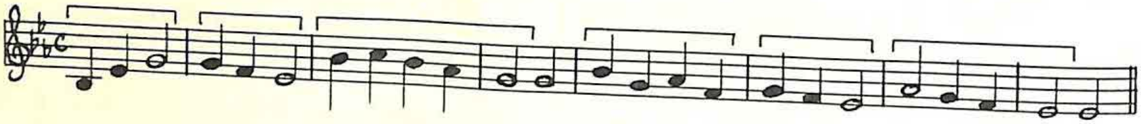


FIGURE 3

THE SONG FOR PHASE II



a) fourth grade, average; b) fourth grade, gifted; c) sixth grade, average; and d) sixth grade, gifted. The subjects within each of these four groups were then assigned at random to either Sequence A or Sequence B, an attempt being made to have equal numbers of subjects in each of the subgroups. This assignment to sequences gave eight groups for each school as shown in Table VII.

Each subject was tested individually, as in Phase I, and the testing consisted of two 15-minute sessions separated by a 48-hour time interval. Before the testing started, each subject was given a brief orientation comparable to the orientation used in Phase I. Subjects in Sequence A were told that during the first session they would try to learn to read five tone groups, or "musical words," and that in two days they would try to learn to read a song which used the same tone groups. Subjects in Sequence B were told that during the first session they would try to learn to read a song and that two days later they were to learn to read five tone groups, or "musical words" which had been used in the song. The orientation also included a short practice session in reading either four tonal configurations or two measures of a song, depending upon the sequence the subject was following.

The procedures for the two tasks were the same regardless of the sequence being followed. The process for learning the five items was identical with the process followed in Phase I and described in detail earlier. There were eight trials of the five items, with the first trial considered as a pre-test, and the items randomized for each successive presentation. Learning the song followed the procedures outlined below:

1. The subject, with the notation of the song before him is given the pitch of the first note of the song and asked to sing as much of it as possible. This served as a pre-test and was used as a basis for determining the amount of learning which took place during the testing.
2. This trial performance was followed by having the subject listen to a performance of the entire song, played at a rate of two quarter notes per second. During the listening the subject was encouraged to follow the notation and, if desired, to sing along with the piano.
3. Following this listening, the subject was asked to sing through the song a second time.
4. Steps 1 and 2 were repeated until the subject had had an opportunity to complete at least eight trials of the song, a trial being defined to include both the singing and listening aspects.

The reader may infer, from these procedures, that a process of rote learning might well be taking place rather than a process of learning to read the notation of the song. To ascertain whether this actually occurred, a check was made of every fourth subject tested. At the conclusion of the testing session, the notation of the song was removed and the subject asked to sing the song from memory. Only a very few subjects were able to proceed as far as the end of the fourth measure; none were able to sing farther than that point; and most subjects were unable to sing more than the first measure and part of the second. These same subjects were, however, able to proceed much farther when using the notation and so it appears reasonable to assume that actual reading rather than rote learning was resulting from the testing.

One significant difference existed between the

## TABLE VII

## THE INITIAL ASSIGNMENT OF THE 151 SUBJECTS PARTICIPATING IN PHASE II

| | 4th Grade | | | | 6th Grade | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sequence "A" | | Sequence "B" | | Sequence "A" | | Sequence "B" | | |
| | Average | Gifted | Average | Gifted | Average | Gifted | Average | Gifted | Totals |
| School A | 10 | 5 | 10 | 5 | 11 | 8 | 10 | 7 | 66 |
| School B | 7 | 4 | 8 | 3 | 4 | 4 | 4 | 4 | 38 |
| School C | 7 | 5 | 7 | 5 | 7 | 6 | 5 | 5 | 47 |
| Totals | 24 | 14 | 25 | 13 | 22 | 18 | 19 | 16 | 151 |

## TABLE VIII

## THE MEAN SCORES FOR THE VISUAL PART OF PHASE I FOR EACH GRADE IN EACH SCHOOL

| | 4th Grade | | 5th Grade | | 6th Grade | | Combined Grades | |
|---|---|---|---|---|---|---|---|---|
| School | N | Mean | N | Mean | N | Mean | N | Mean |
| A | 10 | 4.2 | 10 | 4.9 | 11 | 4.36 | 31 | 4.5 |
| B | 9 | 2.9 | 9 | 2.3 | 10 | 4.1 | 28 | 3.1 |
| C | 11 | 3.5 | 9 | 4.9 | 10 | 5.1 | 30 | 4.4 |
| Total | 30 | 3.53 | 28 | 4.07 | 31 | 4.52 | 89 | 4.0 |

procedures followed in learning the patterns and in learning the song; each pattern was presented individually for both the reading and listening attempts while the song was always presented as a complete entity. In order to minimize this difference, the writer decided that following each reading trial of the song, and prior to the subsequent listening, the reading errors made by the subject would be pointed out and verbally described by the examiner. This decision was based on the analysis of the data from Phase I which showed that many subjects persisted in making the same errors because of an inability to retain the sound of their reading attempt as a basis for comparison with the item when it was played for them. If this held true for items treated individually, it would be more true when the subject was expected to retain the aural memory of an eight-measure song. It was also felt that this assistance resulted in a better attitude on the part of the subjects because they were able to achieve better performances.

For both tasks a criterion of success had been established which consisted of two consecutive perfect performances, or trials, of the task. There seemed no valid reason to continue through all eight trials with subjects who could perform the task after two or three trials.

The equipment used in the test situations was the same as for Phase I and included: a) two tape-recorders; b) one set of test cards for the five items, there being sufficient time to re-arrange them for the next trial; and c) one 5 × 8 index card containing the song, with the notation the same size as used in Phase I. The master tape for the five items was made according to the procedures outlined for the master tape in Phase I; the master tape for the song contained ten successive performances of the song, although eight had been considered sufficient.

Processing the data was carried forward in the same way as for Phase I. The data for both the patterns and the song were transferred to the individual data sheets and scores, in terms of correct responses, were calculated for each subject. Three kinds of scores were used:

1. A "percent correct" score was calculated for each item in the five-item test and for each item in the song, the item being defined as the tonal configuration:

$$\text{"Percent correct"} = \frac{\text{Number of trials item correct}}{\text{Total number of trials}}$$

2. A "mean percent correct" score was calculated for each task. This was a simple mean score:

$$\text{"Mean percent correct"} = \frac{\text{Sum of the "\% correct scores" on each item"}}{\text{Number of items in task}}$$

For both of these scores the unit which served as the basis of the score was the individual item. The first trial was not included in the total number of trials, as explained later.

3. A "rate of learning" score was calculated for each task and represented an attempt to locate a single score which could describe the true net gains made on the task and, at the same time, take into consideration the number of trials needed to make those gains. The formula used for this "RL score" was:

$$RL = \text{true net gain} \times \frac{100}{\text{Total N of trials-1}}$$

The true net gain can be expressed as $\frac{a-b}{c-b}$ in which:

a = Score on final trial, the final trial identified as the first of the two consecutively perfect trials.
b = Score on first trial.
c = Perfect score, the score assigned to any trial in which all items are correct.

The actual gain made by a subject is expressed in the numerator, "a-b;" the total gain possible for the subject is expressed in the denominator, "c-b." The result, which can never exceed 1, is the true net gain.

Since this true net gain does not reflect the number of learning trials taken by the subject, the second part of the formula was used. The numerator, 100, is used so the score may be expressed in larger units. The denominator indicates the actual number of trials during which learning could take place. The first trial for both tasks was considered a pre-test of previous knowledge, since it was assumed that learning the content of the task did not begin until the subject had heard that content. In the case of the song, this occurred after the initial trial; in the case of the tonal configurations, this occurred during the first complete trial but always after the individual item had been attempted. Therefore, this first trial was subtracted from the total number of trials, keeping in mind the definition of the final trial, to give the number of learning trials the subject had.

In Phase I the scoring of the items was done on a "right" or "wrong" basis. It was decided, in view of the nature of the tasks, to use a third alternative for scoring the items in Phase II. This involved giving credit for one-half a correct response if the subject sang the item at the wrong pitch level but retained the same pattern and relative pitches of the original item. For example, if the subject, in reading the item "fa-mi-re-do," did not correctly locate the starting pitch and sang "sol-fa-mi-re" instead, he was given credit for one-half a correct response. This scoring procedure gave data which more accurately described the music reading abilities of the subject since he was not penalized for his inability to sing independently. There are a great many pupils in the elementary grades who can read and interpret music symbols with a high degree of accuracy but who need the support of an accompanying instrument to maintain perfect tonality and intonation. The writer did not, in this exploratory study, wish to discriminate against the child for whom the task was already very difficult.

In order to illustrate the entire scoring procedure, a sample data sheet used for scoring the song is given in Figure 4, and the three scores are indicated as well as the process of deriving the scores. The tonal configurations appearing in the song are indicated with brackets, and the individual notes outside the bracket in measures 4 and 8 are not included in the scoring. A value of eight points is given for each correct tonal configuration yielding a total perfect score of 48 for the six items.

## IV. Results of the Study

### A. Phase I: Visual

Usable data was obtained for 89 of the more than 100 pupils tested in Phase I. Since the subjects had been selected at random from the total fourth-, fifth-, and sixth-grade population in each school, it was highly possible that some subjects would be tested whose data would not be usable. Such subjects would include the following:

1. Pupils who verbally expressed no desire to be tested;
2. Pupils who were unable to match pitches and consequently would not be able to make any singing response to either the visual or aural stimulus;
3. Pupils whose ability to repeat what they heard was so poor that they were unable to sing more than two correct responses out of the ten items;
4. Pupils who became so nervous during the testing that they were unable to continue.

In discussing the data, the term "visual" will be used to refer to the responses made to the visual presentation of the items—the reading aspect of the test. The term "aural" will be used to refer to the responses made to the aural presentation of the items—the listening aspect of the test.

The data for pupils in each grade in each school were first tabulated to ascertain the kinds of scores which had been made. The basis of scoring was to count an item correct if a correct response had been made for any two of the three trials. An item was scored wrong if it was incorrectly sung or if no trial or only a partial trial was made of the item. The score was the sum of the individual items correct, with a maximum score of 10. Table VIII summarizes the mean scores for each grade in each school.

The data shows the mean score for the fourth grade to be lower in each school than the mean scores for the two upper grades. The same difference appears when the three schools are combined by grade level. The data also shows the mean score for the sixth grade to be higher than the mean score for the fifth grade in Schools B and C, but not in School A. When the three schools are combined, however, there is a steady increase in the size of the mean from fourth grade through sixth grade.

A test for the significance of the difference between the mean scores of each grade, combining the three schools, gave F = 1.5068, and it may be concluded that there is no significant difference between the three grades in terms of the data of Phase I (Table IX). The five percent level of significance is used throughout the study. This, of course, has significant implications if intensive testing at these levels continued to support these findings. A test for the significance of the differences between the scores for the fourth and sixth grades, the grades to be studied in Phase II, gave a "t"* of 1.8397 and the differences were considered as not significant.

Table VIII also shows the mean scores for each grade to vary from school to school. For the fourth grade, School A had the highest mean, with schools C and B in descending order. For the fifth grade Schools A and C had identical means, and for the sixth grade School C had the higher mean, with schools A and C in descending order. This would appear to indicate that subjects in each grade in School B consistently performed at a lower level of achievement than subjects from the other two schools. The mean scores for the combined grades in each school shows Schools A and C to have overall means which are very similar

---

*Calculation of "t" used the following formula:

$$\text{“t”} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1 + N_2 - 2}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} \quad \text{in which } \Sigma x^2 = \frac{1}{N}[N\Sigma X^2 - (\Sigma X)^2]$$

FIGURE 4

SAMPLE DATA SHEET, PHASE II, SONG

Mary Jones (Name)
A (Sequence)
Gifted (Group)
A-6 (School and Grade)

| | Item 1 | Item 2[3] | Item 5 | | Item 3 | Item 2[3] | Item 4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| | S D M | M R D | S L S F | M M | S M F R | M R D | F M R | D D | |
| 1 | ✓ | m d $\underline{s}$ | s s m d | d d | r t r t | r t s | m r d | s s | 8 |
| 2 | ✓ | m $\underline{d}$ $\underline{s}$ | s l s m | d d | r d r t | r t s | s f $m^1$ | r d | 12 |
| 3 | ✓ | ✓ | ✓ | ✓ | ✓ | s m d | s f $m^1$ | r d | 36 |
| 4 | ✓ | ✓ | ✓ | ✓ | ✓ | s m d | s f $m^1$ | r d | 36 |
| 5 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | f m d | d d | 40 |
| 6[2] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 48 |
| 7 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| 8 | | | | | | | | | |
| * | 5/5=100% | 4/5=80% | 4/5=80% | | 4/5=80% | 2/5=40% | 2. 5/=50% | | |

[1] Item is counted as half correct and given 4 points because the intervallic relationship is correct even though the pitches are incorrect.

[2] Identified as final trial because it is the first of two consecutively perfect trials.

[3] This item is repeated and accounts for the six items of the task.

*Percent correct score for each item.

$$\text{Mean percent score} = \frac{100+80+80+80+40+50}{6} = \frac{430}{6} = 71.6$$

$$\text{RL Score} = \frac{48-8}{48-8} \times \frac{100}{5} = \frac{40}{40} \times \frac{100}{5} = 20$$

## TABLE IX

## ANALYSIS OF VARIANCE BETWEEN GRADES FOR THE VISUAL PART OF PHASE I

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square |
|---|---|---|---|
| Total | 435.8202 | 88 | |
| Grades | 14.7545 | 2 | 7.3773 |
| Pupils in Grades | 421.0657 | 86 | 4.8961 |
| | | | (F = 1.5068) |

See "Distribution of F" in Palmer O. Johnson's *Statistical Methods in Research* (New York: Prentice-Hall, 1949), pp. 362-65.

## TABLE X

## ANALYSIS OF VARIANCE BETWEEN SCHOOLS FOR THE VISUAL PART OF PHASE I

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square |
|---|---|---|---|
| Total | 435.8202 | 88 | |
| Schools | 33.2830 | 2 | 16.6415 |
| Pupils in Schools | 402.5372 | 86 | 4.6807 |
| | | | (F = 3.5553) |

and, at the same time, more than one point higher than the mean for School B. Analysis of variance yielded an F = 3.5553 which is significant at the five percent level and one may say that a significant difference exists between the three schools.

Further analysis of the data showed the following:

1. There is no significant difference between the mean scores for the combined grades of School A and School C; "t" = .0907.
2. There is a significant difference between the mean scores of School A and School B; "t" = 2.394.
3. There is a significant difference between the mean scores of School C and School B; "t" = 2.2249.

This analysis shows that the subjects from School B did not perform at as high a level as did the subjects from either of the other two schools. In order to accurately identify the reasons for this difference, it would be necessary to consider such factors as: a) the socio-economic background of the subjects; b) the nature of the music program in this school as compared to the other schools; c) the music teacher; and d) the musical interests and music training of the subjects as compared to subjects in other schools.

Examination of the rating sheets completed by the music teacher showed that pupils who were placed in the lowest group in terms of musical ability were always the boys. Not a single girl received so low a rating. The professional literature contains numerous articles dealing with the "boy problem" in music activities at the various grade levels. There is, however, nothing to suggest that the objectives of the music program differentiates between boys and girls and there is no information to indicate that girls possess more native musical ability than do the boys. The conclusion which must be reached is that the girls have a different attitude toward music which makes for better achievement, more interest, and consequently for higher teacher ratings.

The subjects for Phase I had been selected in such a way so as to secure equal numbers of boys and girls at each grade in each school. A summary of the results on the visual part of the test is given in Table XI. This shows the following:

1. For School A, the boys earned higher scores than the girls in both the fourth and sixth grade and these scores were very close together in fifth grade. When the scores for three grades were combined, the boys had a higher mean than did the girls.
2. For School B, the girls earned higher scores than the boys in both the fourth and sixth grades. When the scores for the three grades were combined the girls had a higher mean than did the boys.
3. For School C, the girls consistently did better than the boys.
4. For the three schools combined, only the fourth grade showed little difference between the boys and the girls.
5. For the combined grades in the combined schools, the boys had a mean of 3.67 and the girls had a mean of 4.39.

The "t" test of significance gave a "t" of 1.5457 and it may be concluded that there is no significant difference between the mean scores of the boys and girls when all grades and all schools are combined.

1. Analysis of Reading Errors—A considerable amount of time was spent in making a detailed analysis of the reading errors made on each of the ten items of the test. The purposes of the analysis were: a) to identify any common errors which occurred; b) to attempt to identify the progress shown by the subject for the three trials; and c) to provide necessary data for the item-analysis. The data for the item analysis has been briefly presented. The data regarding the progress of individual pupils is not considered pertinent to the report, and so only the analysis of reading errors will be discussed at this time.

In making this analysis the responses to each item were considered separately for each trial so that a perfect score on all ten items was 30. Multiplying 30 by the number of subjects tested yields the total number of responses possible. Three types of responses were considered: a) right; b) wrong; and c) no trial. A "no trial" response consisted either of no attempt to sing a response to the visual stimulus or failure to reproduce more than one or two notes of the item. Table XII summarizes the responses made by the 89 subjects to each item on the test. The reader is asked to refer to Figure 1 for the notation of these items. Inspection of the totals and means in the table shows the following:

1. Fourth-grade pupils had twice as many "no trial" responses as either of the other two grades. Fifth and sixth grade "no trial" responses are very similar.
2. The mean number of "wrong" responses for sixth grade is slightly lower than that of the other two grades.
3. The mean "right" responses shows improvement from grade to grade although the differences are not statistically significant as was emphasized earlier.
4. The mean "wrong" responses for all grades exceeds the mean "right" responses.
5. Items 2, 3, 4, 7 and 10 had far fewer "right" responses than the other five items on the test.

## TABLE XI

## THE MEAN SCORES FOR THE 43 BOYS AND 46 GIRLS TESTED IN PHASE I

| School | 4th Grade | | | | 5th Grade | | | | 6th Grade | | | | Combined | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Boy | | Girl | | Boy | | Girl | | Boy | | Girl | | Boy | | Girl | |
| | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean |
| A | 5 | 4.6 | 5 | 3.8 | 5 | 4.8 | 5 | 5 | 5 | 4.6 | 6 | 4.2 | 15 | 4.7 | 16 | 4.3 |
| B | 4 | 2.5 | 5 | 3.2 | 4 | 2.5 | 5 | 2.2 | 5 | 3.4 | 5 | 4.8 | 13 | 2.9 | 15 | 3.4 |
| C | 5 | 3.2 | 6 | 3.7 | 5 | 3.0 | 4 | 7.3 | 5 | 4.0 | 5 | 6.2 | 15 | 3.4 | 15 | 5.5 |
| Totals | 14 | 3.5 | 16 | 3.6 | 14 | 3.5 | 14 | 4.6 | 15 | 4.0 | 16 | 5.0 | 43 | 3.6 | 46 | 4.39 |

### TABLE XII

### SUMMARY OF THE RESPONSES MADE BY 89 SUBJECTS TO THE VISUAL PART OF PHASE I

| | 4th Grade | | | | 5th Grade | | | | 6th Grade | | | | Combined Grades | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | N | No Trial | Wrong | Right | N | No Trial | Wrong | Right | N | No Trial | Wrong | Right | N | No Trial | Wrong | Right |
| 1 | 30 | 10 | 32 | 48 | 28 | 3 | 33 | 48 | 31 | 4 | 33 | 56 | 89 | 17 | 98 | 152 |
| 2 | 30 | 8 | 69 | 13 | 28 | 7 | 62 | 15 | 31 | 4 | 80 | 9 | 89 | 19 | 211 | 37 |
| 3 | 30 | 15 | 69 | 6 | 28 | 10 | 69 | 5 | 31 | 6 | 80 | 7 | 89 | 31 | 218 | 18 |
| 4 | 30 | 14 | 62 | 14 | 28 | 4 | 57 | 23 | 31 | 3 | 57 | 33 | 89 | 21 | 176 | 70 |
| 5 | 30 | 4 | 27 | 59 | 28 | 0 | 31 | 53 | 31 | 4 | 27 | 62 | 89 | 8 | 85 | 174 |
| 6 | 30 | 6 | 55 | 29 | 28 | 3 | 50 | 31 | 31 | 3 | 47 | 43 | 89 | 12 | 152 | 103 |
| 7 | 30 | 4 | 63 | 23 | 28 | 3 | 62 | 19 | 31 | 3 | 61 | 29 | 89 | 10 | 186 | 71 |
| 8 | 30 | 4 | 34 | 52 | 28 | 4 | 29 | 51 | 31 | 2 | 25 | 66 | 89 | 10 | 88 | 169 |
| 9 | 30 | 4 | 17 | 69 | 28 | 2 | 18 | 64 | 31 | 2 | 23 | 68 | 89 | 8 | 58 | 201 |
| 10 | 30 | 6 | 63 | 21 | 28 | 1 | 61 | 22 | 31 | 1 | 58 | 34 | 89 | 8 | 182 | 77 |
| Totals | 30 | 75 | 491 | 334 | 28 | 37 | 472 | 331 | 31 | 32 | 491 | 407 | 89 | 144 | 1454 | 1072 |
| Mean | | 2.5 | 16.4 | 11.1 | | 1.3 | 16.8 | 11.8 | | 1.0 | 15.8 | 13.1 | | 1.6 | 16.3 | 12.04 |

The scope of this report is such that a detailed discussion of all reading errors for each item would prove impractical. The writer proposes, therefore, to present a brief summary of the errors made most frequently by the 89 subjects, using tonic sol-fa syllables for identification. A dash below the syllable indicates that the pitch is lower than the preceeding tone; a dash above indicates that the pitch is higher.

Item 1: (F M̲ R̲ D̲)

Within the 98 errors made by the combined grades, the following appeared most frequently:

| | | |
|---|---|---|
| 1) | FR̲T̲S̲ | 10 |
| 2) | FR̲T̲F̲ | 5 |
| 3) | FR̲M̅D̲ | 5 |
| 4) | FS̅L̅F̲ | 5 |
| 5) | F̅S̅L̅T̅ | 5 |
| 6) | FM̲e̲R̲D̲ | 4 |
| 7) | FR̲D̲T̲e̲ | 4 |
| 8) | SM̲D̲S̲ | 5 |
| 9) | SF̲M̲R̲ | 6 |
| | Total | 49 |

The remaining 49 errors had frequency counts of 1 to 3, 25 errors having a frequenay count of only 1. It is interesting to note that Error 5 is an exact inversion of the pattern and Error 4 a partial inversion; that errors 1, 2, and 8 are descending chord figures; and that Errors 8 and 9 began on the wrong pitch despite the fact that the actual starting pitch was sounded for two seconds. In fact, 27 of the 98 errors included failure to begin on the correct pitch. The scores for the three grades are very similar although the fourth grade had more "no trial" responses.

Item 2: (D L̲ S̲)

Within the 211 errors of the combined grades, these appeared most frequently:

| | | |
|---|---|---|
| 1) | DT̲L̲ | 71 |
| 2) | DS̲M̲ | 44 |
| 3) | DL̲F̲ | 26 |
| 4) | DS̲L̲ | 8 |
| 5) | DS̲F̲ | 9 |
| 6) | DS̲D̲ | 8 |
| 7) | DT̲S̲ | 6 |
| 8) | DT̲e̲L̲ | 4 |
| | Total | 177 |

The eight errors account for 177, or 84 percent of the total wrong responses. Of remaining 34 errors, 17 had a frequency of 1, four had a frequency of 2, and three had a frequency of 3. It is interesting to note that Errors 1, 2, and 3 were the most common. Children do not appear to notice the interval of the minor third followed by the major second since they sing the pattern either as a partial scale or as a complete chord figure. There were only 37 correct responses out of a possible total of 267. It is of interest to note that the sixth grade had more "wrong" responses and fewer "right" responses on this item than either of the other two grades. Only 17 errors were caused by failure to start on the correct pitch.

Item 3: (S M̲ D̲ R̅ T̲ D̅)

The errors appearing most frequently in the total of 218 were:

| | | |
|---|---|---|
| 1) | SM̲D̲M̅D̲M̅ | 14 |
| 2) | SF̲M̲F̅M̲R̲ | 13 |
| 3) | SF̲M̲S̅F̲M̲ | 13 |
| 4) | SF̲M̲F̅R̲M̅ | 10 |
| 5) | SF̲M̲R̲D̲T̲ | 7 |
| 6) | SF̲M̲F̅S̅F̲ | 6 |
| 7) | SM̲D̲R̅S̲D̅ | 6 |
| 8) | SM̲D̲M̅R̲D̲ | 5 |
| 9) | SM̲D̲M̅S̲D̅ | 5 |
| | Total | 79 |

These nine errors account for only 79, or 3.6 percent of all errors with most of the remaining 139 having frequency counts of only 1 or 2. This item was the most complex because it included six tones and was very disjunct, and this accounts for the fact that it had the fewest number of correct responses of any item. The same held true for the aural responses to the item; it again had the lowest number of correct responses.

Errors 1, 7, 8 and 9 show a partially correct response being made, presumably because the first part of the pattern was familiar and could be remembered better than the last three notes. Error 5 is a simplification of the pattern so it becomes a scale while the other errors show a tendency to repeat a 3-tone pattern. Failure to begin on the starting pitch accounts for only 32 of the errors.

Item 4: (D S̲ L̅ T̅ D̅)

This item began with a descending interval of a fourth and continued with an ascending scale pattern which returned to the starting tone. The interval is a difficult one to sing and this is evident in the number of errors. The 9 errors most frequently made accounted for 100, or 57 percent of the total "wrong" responses.

| | | | | | |
|---|---|---|---|---|---|
| 1) | $\mathrm{D\underline{TLSF}}$ | 32 | 4) | $\mathrm{D\underline{D}\overline{RMF}}$ | 13 |
| 2) | $\mathrm{D\underline{TL}\overline{TD}}$ | 19 | 5) | $\mathrm{D\underline{M}\overline{FSL}}$ | 8 |
| 3) | $\mathrm{D\underline{L}\overline{TDR}}$ | 16 | 6) | $\mathrm{D\underline{LSFM}}$ | 7 |
| | | | 7) | $\mathrm{D\underline{F}\overline{SL}T}$ | 5 |
| | | | | Total | 100 |

Errors 1 and 2 completely ignore the beginning interval in the item either because it was not perceived or because the subjects were unable to produce any interval larger than a second. Error 6 under estimates the size of the beginning interval but does acknowledge its presence, and then ignores the ascending scale and transforms it into a descending scale. The remaining errors all show recognition of the general shape of the item, a skip down followed by an ascending scale, but fail to get the interval correct. This characteristic was also true for most of the remaining 76 errors which had frequencies of 1, 2 or 3. The sixth grade had considerably more "right" responses to the item than the other two grades. Only 13 errors were the result of an inability to match the starting tone which was given.

### Item 5: $(\mathrm{D\overline{M}\,\overline{S}})$

This is one of the more familiar items on the test, a fact reflected in the large number of "right" responses by all grades. Note the way in which the pupils in the fourth and sixth grades duplicate each others performance. This item had only 85 "wrong" responses, a record exceeded only by item 9, and one error, "DRM" had a frequency of 53, or 62 percent of the total. In fact, there were only 18 different errors counted for this item, the other 17 having frequencies of 1 or 2. The tonic chord is often sung as a 3-note scale while DRM is often sung as the tonic chord. This suggests that subjects do not really read the notation but more often guess at it. A total of 13 errors resulted from failure to begin on the correct pitch.

### Item 6: $(\mathrm{S\overline{L}\underline{S}\,\underline{F}\,\underline{M}})$

This combination of neighboring tones and a partial descending scale was easier for the sixth grade than for fourth and fifth, whose scores are similar. A total of 62 different errors were counted and the 7 most frequently made accounted for 75, or 49 percent of the total number of "wrong" responses. These are:

| | | | | | |
|---|---|---|---|---|---|
| 1) | $\mathrm{S\overline{L}\underline{SMD}}$ | 20 | 4) | $\mathrm{S\underline{M}\overline{F}\underline{RD}}$ | 9 |
| 2) | $\mathrm{S\underline{FMRD}}$ | 15 | 5) | $\mathrm{S\overline{L}\underline{FMR}}$ | 8 |
| 3) | $\mathrm{S\overline{D}\underline{SMD}}$ | 12 | 6) | $\mathrm{S\overline{L}\underline{SFR}}$ | 6 |
| | | | 7) | $\mathrm{S\overline{L}\underline{SMR}}$ | 5 |
| | | | | Total | 75 |

Errors 1, 6 and 7 show a partial success in reading the item, while Error 2 results again from attempting to use a familiar response to notation which is not carefully observed. Errors 3, 4 and 5 represent inaccurate responses which were strongly influenced by the aural presentation of the item in the trial which preceded this particular item. Only 12 errors were identified as failure to begin on the correct note.

### Item 7: $(\mathrm{\underline{S}\overline{D}M})$

This item, an inversion of the more familiar item 5, contains within its first two notes one of the most common tone patterns identified in songs. It ranked first in the analysis made by the writer and Smiley's analysis, and is invariably found in songs which have anacrusis, or "pick-up note." Despite this, the fact that it begins with an ascending fourth presents a definite reading problem to children in all the grades. Many children experienced difficulty in singing the first tone because they were unable to relax sufficiently for the note to sound. This difficulty did not appear in the analysis of aural responses to the same item although for the reading response it apparently was a factor since 47, or 25 percent of the total errors were the result of an incorrect beginning pitch. The common errors identified for this item are:

| | | | | | |
|---|---|---|---|---|---|
| 1) | $\mathrm{\underline{S}\overline{TR}}$ | 49 | 5) | $\mathrm{S\overline{TD}}$ | 8 |
| 2) | $\mathrm{\underline{S}\overline{LT}}$ | 36 | 6) | $\mathrm{S\overline{RF}}$ | 6 |
| 3) | $\mathrm{D\overline{RM}}$ | 10 | 7) | $\mathrm{S\overline{L}S}$ | 5 |
| 4) | $\mathrm{S\overline{LD}}$ | 9 | 8) | $\mathrm{T\overline{RF}}$ | 4 |
| | | | 9) | $\mathrm{L\overline{DM}}$ | 4 |
| | | | 10) | $\mathrm{D\overline{MS}}$ | 4 |
| | | | | Total | 135 |

The 10 errors account for 73 percent of the total "wrong" responses and Errors 1 and 2 account for 46 percent of this same total. The tendency to reduce the interval of a fourth to an interval of a third and proceed to sing a tonic chord is evident in Error 1; the tendency to revert to a scale pattern whenever the reader is in doubt is shown in Error 2.

Item 8: (MRD)

This is one of four items for which the number of "right" responses exceeds the number of "wrong" responses. The fourth- and fifth-grade pupils perform almost the same way, with the sixth grade showing better results. Only 17 of the 88 errors resulted from beginning on the wrong pitch. The familiarity of this pattern is established by the scores. The common errors are:

| | | | |
|---|---|---|---|
| 1) | MDS | 24 | |
| 2) | MF̅S | 10 | |
| 3) | SMD | 6 | |
| 4) | MDL | 5 | |
| 5) | MF̅R | 5 | |
| 6) | SFM | 4 | |
| | Total | 54 | |

These errors account for 61 percent of the total and it is interesting to note the frequency with which this simple scale pattern becomes a descending tonic chord, Error 1 has a frequency of 24, or 27 percent of the total number of errors. The reader is referred to the analysis of Item 5 which shows parallel results.

Item 9: (LSFMRD)

This six-note descending scale proved to be the easiest item of the ten, and all grades appeared to perform with equal ability. There were no errors having a frequency greater than five although a total of 37 different errors were counted. Failure to begin on the starting tone did account for 20, or 35 percent of the total errors made.

Item 10: (SMF̅R)

The disjunct character of this item made it more difficult than one would expect since a sequence of descending thirds occurs quite often in songs. Again, very little difference exists between the fourth and fifth grades while the sixth grade achieves slightly better scores. A total of 59 different errors were tabulated and all but the four given below had frequency counts of six or less:

| | | |
|---|---|---|
| 1) | SFMR | 27 |
| 2) | SMS̅M | 14 |
| 3) | SMF̅D | 14 |
| 4) | SRM̅D | 10 |
| | Total | 65 |

Error 1 again represents this phenomenon of substituting a familiar tonal configuration for an unfamiliar one, possibly the result of carelessness but more often resulting from the subject's inability to "hear" the pattern before it is sung. Only nine errors of the 182 were caused by an incorrect starting pitch.

2. Summary of the Findings—This detailed analysis of reading errors emphasizes several important points. The total of 1454 "wrong" and 144 "no trial" responses as compared with 1072 "right" responses shows that children in the elementary grades experience considerable difficulty in reading tonal configurations which are commonly found in the songs children sing.

1. During the testing period, the writer found that most children read the item very slowly, a single note at a time. This is not apparent in the data because there is no way of indicating this kind of performance other than to listen to the tape-recorded responses.
2. Although children experienced little difficulty in making a response to the aural stimulus they were seemingly unable to "make the voice go the way the notes do," as several subjects pointed out. This implies that skill in reading music can be developed only if the child is able to "hear" the item silently before he sings it. Developing this internal "hearing" presents many problems including an adequate identification of effective teaching procedures.
3. Children are usually aware of the general shape or direction of the tonal configurations but fail to acquire the ability to accurately identify the internal changes which make one pattern different from another even though the two patterns appear quite similar.
4. When children are asked to read tonal configurations which are not immediately recognized as something they know, one of two types of responses generally result: 1) changing the pattern so that it becomes a familiar pattern which is very similar to the stimulus, or 2) just guessing that the response being made is something like the stimulus without having any sound basis upon which to judge the accuracy of their response. The total number of different errors which the children made on the 10 items in evidence of the second kind of response. Excluding the "no trial" responses, 489 different errors were made in accumulating the 1454 "wrong" responses, giving an average frequency of 2.97 per error. Evidence of the first type of response is indicated by the fact that 61, or 12.47 percent of the 489 different errors

had a total frequency count of 787, or 54 percent of the 1454 "wrong" responses.

5. The data from Phase I shows comparatively little difference between grades in terms of general ability to perform the task. Part of this might be due to the fact that three trials of the 10 items did not give the subjects sufficient opportunity to learn the items. The question arises as to whether more trials would have shown significant differences between the three grades. Since most music reading activities begin at the third-grade level, this means that sixth-grade children have two additional years of experience with reading music than do fourth-grade children. If there is no significant difference between the achievement of the two grades, and the data from Phase I shows that this is true in terms of this particular investigation, some highly interesting speculations may be made as to why this is true.

## B. Phase I: Aural

Paralleling the analysis of reading errors was an analysis of the errors made by the subjects as they responded to the aural presentation of each of ten items. It is generally agreed by music educators that music reading ability depends a great deal upon developing the ability to perceive music aurally. This "ear vocabulary" of musical sounds establishes the basis for music reading activities, because the child is then able to work with familiar concepts and materials.

The data on the singing responses to the aural presentation of the item was processed in the same way as the data for the visual part of the study. The basis of scoring was again to count an item correct if a correct response had been made for any two of the three trials. An item was counted as wrong if it was incorrectly sung or if no trial or only a partial trial had been made on the item. The score was the sum of the individual items correct, with a maximum of 10. Table XIII summarizes the mean scores for each grade in each school.

The data shows that fourth- and fifth-grade subjects, in each school and for the combined schools, had lower mean scores than the sixth-grade subjects. Only in School B was there a marked difference between the means of fourth- and fith-grade subjects. These results are similar to the results on the visual part of the test in terms of the relative achievement of each grade. The reader may wish to compare these means with the means on the visual part presented in Table VIII. Such a comparison shows that pupils in each grade earned a score on the aural part of the test that was approximately four points higher than the scores for the visual part. In other words, they were able to double their achievement when an imitative, rather than a reading response was required.

A test for the significance of the differences between the mean scores of each grade, combining the three schools, gave F = 1.6297 and it may be concluded that there is no significant difference between the three grades in terms of this data (Table XIV).

This may well indicate that the ability to respond to an aural stimulus is already developed by the fourth-grade level and does not change significantly after that time.

Table XIII shows the mean scores for each grade in School A and C to be very similar but that subjects in School B consistently performed at a lower level of achievement, especially when the fifth grades are compared. However, analysis of variance between the schools gave F = 2.7376 and it may be concluded that no significant difference exists between the three schools when subjects are asked to respond to an aural stimulus (Table XV).

There is no reason to suspect that boys differ significantly from girls in ability to respond to an aural stimulus except the difference which may be due to the general inclination of the boy to avoid using his singing voice. The combined mean score for the 43 boys tested was 8.023 and for the girls it was 8.587. The "t" test of significance gave "t" = 1.6386 and it may be concluded that there is no significant difference between the mean scores of boys and girls.

These data would tend to support the point-of-view that auditory perception, as measured by a singing response to an aural stimulus, is already developed by the time a child reaches the fourth grade and does not significantly change as a result of the additional musical experiences in the upper elementary grades.

1. Analysis of Singing Errors—A detailed analysis of the singing errors made on each of the ten items was carried forward following the procedures used in making the analysis of reading errors. The reader has observed the fact that the subjects had considerably higher mean scores on the aural part of the test than they did on the visual part. The writer was particularly interested in whether subjects with high scores on the aural part of the test had also obtained high scores on the visual part. The product-moment correlations were computed and resulted in the following:

| | |
|---|---|
| 1. Sixth grade, | r = .64 |
| 2. Fifth grade, | r = .70 |
| 3. Fourth grade, | r = .58 |
| 4. Combined grades, | r = .64 |

The writer is unable, at this time, to determine whether these could be considered high, medium, or low correlations in terms of the two skills being considered. It is of interest to note that fifth-

## TABLE XIII

## THE MEAN SCORES FOR THE AURAL PART OF PHASE I FOR EACH GRADE IN EACH SCHOOL

| School | 4th | | 5th | | 6th | | Combined | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | N | Mean | N | Mean | N | Mean |
| A | 10 | 8.5 | 10 | 8.5 | 11 | 8.8 | 31 | 8.6 |
| B | 9 | 7.8 | 9 | 7.0 | 10 | 8.4 | 28 | 7.8 |
| C | 11 | 8.3 | 9 | 8.4 | 10 | 8.9 | 30 | 8.5 |
| Total | 30 | 8.2 | 28 | 8.0 | 31 | 8.7 | 89 | 8.3 |

## TABLE XIV

## ANALYSIS OF VARIANCE BETWEEN GRADES FOR THE AURAL PART OF PHASE I

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square |
|---|---|---|---|
| Total | 219.1910 | 88 | |
| Grades | 8.0039 | 2 | 4.0020 |
| Pupils in Grades | 211.1871 | 86 | 2.4557 |
| | | | (F = 1.6297) |

## TABLE XV

## ANALYSIS OF VARIANCE BETWEEN SCHOOLS FOR THE AURAL PART OF PHASE I

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square |
|---|---|---|---|
| Total | 219.1910 | 88 | |
| Schools | 13.1195 | 2 | 6.5598 |
| Pupils in Schools | 206.0715 | 86 | 2.3962 |
| | | | (F = 2.7376) |

grade children are apparently more consistent than children in the other two grades, with the fourth grade not maintaining as high a straight-line "relationship." This is a problem which needs further investigation before any sound conclusions can be drawn.

The data presented in Table XVI summarizes the responses of the 89 subjects, these responses again being considered separately for each trial so that a perfect score on all ten items is 30. The table shows the following:

1. There were very few "no trial" responses, only 16 for the three grades, with the sixth grade having only one.
2. Sixth-grade pupils had fewer "wrong" responses, a mean of 4.1, while fourth and fifth grades had means of 5.67 and 6.1 respectively.
3. The mean "right" responses show the sixth grade to be slightly better than the other two grades.
4. The mean "right" responses for all grades is almost five times greater than the mean "wrong" responses.
5. Only item 3 shows the "wrong" responses to exceed the "right" responses, a result attributed to the length and complexity of the item.

Since the reader may be interested in comparing the visual and aural parts of the test, Table XVII gives a brief summary of the pertinent information. This clearly points out the comparative difficulty of the two aspects of the test.

The writer does not feel that any specific purpose would be served by presenting a detailed, item-by-item analysis of the singing errors made by the subjects. Inspection of the individual data sheets showed that a total of 246 different errors were made in accumulating the 466 "wrong" responses. It will be recalled that analysis of the reading errors for this test gave 61 common errors which had a frequency count of four or more and accounted for 54 percent of the total number of "wrong" responses. For the aural responses, only 22 such common errors were noted and these accounted for 36 percent of the total number of "wrong" responses. One reason for the lack of common errors lies in the difference between the two tasks. For the visual part of the study, with the subject singing the starting pitch immediately after it had been given, failure to begin on the correct pitch accounted for 1.4 percent of the total errors; for the aural part of the study it accounts for 39.3 percent of the total errors. This inability to locate the correct starting pitch for items presented aurally is probably due to the fact that the subject had to retain this pitch during the entire aural presentation, something which many of the pupils were unable to do.

2. Summary of Findings—The analysis of singing errors which has been briefly presented shows that aural imitation of an item is considerably easier for children than making responses to visual stimuli. While 54.5 percent of the responses to the visual presentation of the items were "wrong" responses, only 17.5 percent of the responses to aural presentations were "wrong". This again emphasizes the difference in the difficulty of the two kinds of responses. The "wrong" responses to the aural presentation of the items showed few common errors and this suggests that guessing is taking place. It appears unlikely that music reading skill can be developed by continuing to emphasize rote-learning and imitative singing in the upper grades. The data show that little difference exists between grades in terms of this auditory perception. The major problem centers around the methods used to help the child relate this auditory perception to the visual stimuli so the latter are meaningful.

### C. Phase II

Originally 151 subjects had been chosen for this aspect of the study but several of these were not available for testing and so data was secured for 138 fourth- and sixth-grade pupils. Table XVIII shows the distribution of these subjects according to school, grade, and the sequence to which they had been assigned. The procedures for identifying the "average" and "musically gifted" subjects have been described earlier in the report.

Certain additional information was obtained for each subject and included: 1) age, in months, through March 1, 1959; 2) I.Q., as determined by the California Test of Mental Maturity, Form S for Elementary Grades; and 3) the kind of musical instrument played and the length of time it had been studied.

Table XIX summarizes the information regarding age and I.Q., of the subjects, with the figures given being the means for each group in each school. This shows relatively little variation in age for each group in the three schools. There is a greater difference between the means of the various groups in terms of I.Q. The "gifted" group in each grade has a higher mean I.Q., than the "average" group in the same grade and this difference is maintained for each of the three schools. It is of interest to note that the mean I.Q., of subjects in School C is higher than in either of the other two schools, with School A slightly lower than School B. Unfortunately, it has not been possible, within the scope of this study, to ascertain either the significance of the differences or the possible effect I.Q., might have upon the subject's performance in the tests of Phase II. Research on the topic tends to show that a high level of native musical capacity is not necessarily accom-

## TABLE XVI

### SUMMARY OF THE RESPONSES MADE BY 89 SUBJECTS TO THE AURAL PART OF PHASE I

| | 4th Grade | | | | 5th Grade | | | | 6th Grade | | | | Combined Grades | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | N | No Trial | Wrong | Right | N | No Trial | Wrong | Right | N | No Trial | Wrong | Right | N | No Trial | Wrong | Right |
| 1 | 30 | 0 | 8 | 82 | 28 | 1 | 13 | 70 | 31 | 0 | 5 | 88 | 89 | 1 | 26 | 240 |
| 2 | 30 | 2 | 8 | 80 | 28 | 1 | 21 | 62 | 31 | 0 | 13 | 80 | 89 | 3 | 42 | 222 |
| 3 | 30 | 1 | 60 | 29 | 28 | 2 | 55 | 27 | 31 | 0 | 56 | 37 | 89 | 3 | 171 | 93 |
| 4 | 30 | 1 | 16 | 73 | 28 | 2 | 14 | 68 | 31 | 0 | 10 | 83 | 89 | 3 | 40 | 224 |
| 5 | 30 | 1 | 6 | 83 | 28 | 1 | 3 | 80 | 31 | 0 | 2 | 91 | 89 | 2 | 11 | 254 |
| 6 | 30 | 1 | 15 | 74 | 28 | 0 | 17 | 67 | 31 | 1 | 9 | 83 | 89 | 2 | 41 | 224 |
| 7 | 30 | 0 | 10 | 80 | 28 | 0 | 7 | 77 | 31 | 0 | 9 | 84 | 89 | 0 | 26 | 241 |
| 8 | 30 | 0 | 4 | 86 | 28 | 0 | 8 | 76 | 31 | 0 | 4 | 89 | 89 | 0 | 16 | 251 |
| 9 | 30 | 2 | 28 | 60 | 28 | 0 | 25 | 59 | 31 | 0 | 15 | 78 | 89 | 2 | 68 | 197 |
| 10 | 30 | 0 | 15 | 75 | 28 | 0 | 7 | 77 | 31 | 0 | 3 | 90 | 89 | 0 | 25 | 242 |
| Totals | 30 | 8 | 170 | 722 | 28 | 7 | 170 | 663 | 31 | 1 | 126 | 803 | 89 | 16 | 466 | 2188 |
| Mean | | .03 | 5.67 | 24.1 | | .03 | 6.1 | 23.7 | | .003 | 4.1 | 26.0 | | .18 | 5.2 | 24.6 |

TABLE XVII

SUMMARY OF THE RESPONSES MADE BY 89 SUBJECTS TO BOTH THE VISUAL AND AURAL PARTS OF PHASE I

| | No Trial | | | | Wrong | | | | Right | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Visual | | Aural | | Visual | | Aural | | Visual | | Aural | |
| Grade | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean |
| 4th | 75 | 2.5 | 8 | .03 | 491 | 16.4 | 170 | 5.67 | 334 | 11.1 | 722 | 24.1 |
| 5th | 37 | 1.3 | 7 | .03 | 472 | 16.8 | 170 | 6.1 | 331 | 11.8 | 663 | 23.7 |
| 6th | 32 | 1.0 | 1 | .003 | 491 | 15.8 | 126 | 4.1 | 407 | 13.1 | 803 | 26.0 |
| Combined | 144 | 1.6 | 16 | .02 | 1454 | 16.3 | 466 | 5.2 | 1072 | 12.0 | 2188 | 24.6 |

Note: The total number of possible responses is 2670, obtained by multiplying the number of responses on the three trials (30), by the number of subjects (89).

TABLE XVIII

THE DISTRIBUTION OF THE 138 SUBJECTS TESTED IN PHASE II

| | 4th Grade | | | | 6th Grade | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sequence "A" | | Sequence "B" | | Sequence "A" | | Sequence "B" | | |
| School | Average | Gifted | Average | Gifted | Average | Gifted | Average | Gifted | Total |
| A | 7 | 6 | 8 | 4 | 14 | 9 | 9 | 6 | 63 |
| B | 4 | 4 | 5 | 3 | 4 | 5 | 5 | 3 | 33 |
| C | 5 | 6 | 6 | 4 | 4 | 8 | 4 | 5 | 42 |
| Totals | 16 | 16 | 19 | 11 | 22 | 22 | 18 | 14 | 138 |

## TABLE XIX

## SUMMARY OF THE AGES AND I.Q.'s OF THE 138 SUBJECTS IN PHASE II

| | Age | | | | I.Q. | | | |
|---|---|---|---|---|---|---|---|---|
| | 4 | | 6 | | 4 | | 6 | |
| School | Average | Gifted | Average | Gifted | Average | Gifted | Average | Gifted |
| A | 118.9 | 119.1 | 141.5 | 141.7 | 107.9 | 111.5 | 110.1 | 114.1 |
| B | 116.4 | 116.6 | 144.9 | 140.6 | 114.2 | 118.0 | 111.2 | 120.1 |
| C | 117.4 | 114.9 | 142.0 | 142.0 | 120.3 | 127.6 | 116.3 | 123.1 |
| CS* | 117.8 | 116.9 | 142.5 | 141.6 | 113.2 | 119.2 | 111.8 | 118.5 |
| CAG** | 117.4 | | 142.0 | | 115.8 | | 115.2 | |

* "CS" = Combined Schools
**"CAG" = Combined Average and Gifted

panied by a similar superiority of intelligence. There is little information available regarding the possible effect I. Q., might have upon musical achievement although it may be assumed that these two are more closely related than I. Q., and musical capacity. The extent of this latter relationship needs to be determined by careful research.

There is reason to believe that experience in playing a musical instrument results in a higher level of music reading ability because instrumental performance tends to crystallize the concepts of pitch and tonal relationships as expressed by musical notation. This information was secured for each subject participating in Phase II and is summarized in Table XXa. The data includes only those subjects reporting instrumental training of 12 months or more. The writer assumed that a shorter period of instruction would probably have little effect upon the ability to read music. The table clearly shows that more "gifted" subjects had such training than "average" subjects. The total N of 16 fourth-grade subjects accounts for 25.8 percent of the total fourth-grade sample; the 37 sixth-grade subjects are 48.7 percent of the sixth-grade sample. When the total group of 138 is considered, the data shows that 53 subjects or 38.4 percent had at least 12 months of instruction. The differences between the three schools, in terms of the sample, may be more clearly shown when the data is expressed as "percent of the total sample," as in Table XXb.

1. Test Results: Tonal Configurations—The results of the testing in Phase II will be presented in terms of each task, rather than in terms of each sequence of tasks, in order to simplify the presentation. The two tasks were: 1) to learn five tonal configurations within a 15-minute testing period; and 2) to learn a short song constructed from these five tonal configurations within a similar 15-minute testing period. For Sequence A, the tonal configurations were first, followed in two days by the song; for Sequence B, the tasks were reversed.

Two kinds of scores will be used to present the results; the "percent correct" scores, and the "rate-of-learning" score. These have been described earlier in the report and it will be remembered that "percent correct" score for each item was calculated by summing the number of correct responses to the item and dividing by the total number of trials for the item; summing the "percent correct" scores for each item and dividing by the number of items gave the "mean percent correct" score for the total task.

One of the first steps taken was to ascertain the reliability of the tests used in Phase II. The subjects in each of the two sequences at each grade level were divided into two equal groups, assignment to groups being made at random. The "average" and "gifted" pupils were combined, as were the three schools. If the differences between the means for each of these two groups was not significant it could be assumed that the tests were reliable, and that any other differences were due to factors other than the tests. Table XXI presents the data for both the learning of tonal configurations and the learning of the song. The "t" values for the split groups in each grade and for each sequence show that no real difference exists between the means and the tests may be considered reliable for this particular sample. It should be pointed out that the scores on the song represent the percentage of correct responses to the tonal configurations within the song.

Analysis of the "mean percent correct" scores for tonal configurations that were presented as a separate task is summarized in Table XXII. The highest possible score is 100, indicating that all five items were correct on every trial taken by the subject. The reader is reminded that for Sequence A this was the initial task, but for Sequence B it was the second task and was performed after the subject had had practice in learning the song. The higher means for average and gifted subjects in each grade who were tested in Sequence B are probably due to this practice which they received. Examination of the individual tabulations shows that a total of 12 fourth-grade pupils and 20 sixth-grade pupils had scores of 100. This represents 23.2 percent of the total sample of 138. More gifted subjects, 10 fourth grade and 18 sixth grade, had perfect scores and it is of interest to note that at the sixth-grade level, 13 of the 20 perfect scores were in Sequence A. The fourth grade showed 9 perfect scores for Sequence B, and only one in Sequence A. It would seem that the practice effect for Sequence B is more evident at the fourth-grade level than the sixth-grade level. It is also possible that the gifted sixth-grade subjects in Sequence A were more competent than their peers in Sequence B.

The data in Table XXII shows the relative superiority of the gifted subjects as compared with average subjects. For each sequence in each grade, the gifted group has a considerably higher mean, a performance which could have been predicted without making the study. Of more interest and of greater significance is the comparison of each group in Sequence A with its parallel group in Sequence B to ascertain the degree to which practice on a related task influences test scores. The "t" tests are shown at the top of

These data appear to support the following conclusions:

1. Practice is more necessary for average fourth-grade pupils than any other group. This group needs to develop an understanding of the concepts involved in reading musical notation.

## TABLE XXa

## MEAN NUMBER OF MONTHS OF INSTRUMENTAL INSTRUCTION FOR 53 SUBJECTS IN PHASE II

| | 4th Grade | | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|
| | Average | | Gifted | | Average | | Gifted | |
| School | N | Mean | N | Mean | N | Mean | N | Mean |
| A | 0 | 0 | 6 | 23 | 4 | 15 | 11 | 32 |
| B | 2 | 24 | 4 | 21 | 3 | 19 | 5 | 39.6 |
| C | 1 | 24 | 3 | 44 | 4 | 31.5 | 10 | 38.4 |
| Totals | 3 | 24 | 13 | 27.2 | 11 | 22.2 | 26 | 35.9 |
| Grand Total | N = 16 | | Mean = 26.6 | | N = 37 | | Mean = 31.8 | |

## TABLE XXb

## PERCENT OF SAMPLE REPORTING INSTRUMENTAL INSTRUCTION

| | 4th Grade | | | 6th Grade | | | 4th and 6th Grades | | |
|---|---|---|---|---|---|---|---|---|---|
| School | N | n | % | N | n | % | N | n | % |
| A | 25 | 6 | 24.0 | 38 | 15 | 38.5 | 63 | 21 | 33.3 |
| B | 16 | 6 | 37.5 | 17 | 8 | 47.1 | 33 | 14 | 42.4 |
| C | 21 | 4 | 19.1 | 21 | 14 | 66.7 | 42 | 18 | 42.9 |

N = Total number of subjects for each school in each grade.
n = Number of subjects reporting instrumental instruction.
% = n/N

## TABLE XXI

"MEAN PERCENT CORRECT" SCORES FOR TWO EQUAL GROUPS

(A) Tonal Configurations

| | 4th Grade | | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|
| | Sequence A | | Sequence B | | Sequence A | | Sequence B | |
| | I | II | I | II | I | II | I | II |
| N | 16 | 16 | 15 | 15 | 22 | 22 | 16 | 16 |
| ΣX | 1004 | 935 | 1109 | 1147 | 1703 | 1489 | 1254 | 1140 |
| M | 62.75 | 58.44 | 73.93 | 76.47 | 77.27 | 67.68 | 78.38 | 71.25 |
| $\Sigma X^2$ | 78190 | 65945 | 90505 | 96113 | 145175 | 113671 | 111270 | 91666 |
| | "t" = .4103 | | "t" = .2829 | | "t" = 1.2719 | | "t" = .7217 | |

(B) Song

| | 4th Grade | | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|
| | Sequence A | | Sequence B | | Sequence A | | Sequence B | |
| | I | II | I | II | I | II | I | II |
| N | 16 | 16 | 15 | 15 | 22 | 22 | 16 | 16 |
| ΣX | 985 | 988 | 547 | 675 | 1831 | 1647 | 831 | 855 |
| M | 61.56 | 61.75 | 36.47 | 45.0 | 83.28 | 74.86 | 51.94 | 53.44 |
| $\Sigma X^2$ | 71867 | 70734 | 28479 | 40191 | 156341 | 132793 | 53249 | 54849 |
| | "t" = .0203 | | "t" = .9123 | | "t" = 1.5509 | | "t" = .1674 | |

## TABLE XXII

### "MEAN PERCENT CORRECT" SCORES FOR AVERAGE AND GIFTED SUBJECTS ON THE TONAL CONFIGURATIONS

| | 4th Grade | | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|
| | Sequence A | | Sequence B | | Sequence A | | Sequence B | |
| | Average | Gifted | Average | Gifted | Average | Gifted | Average | Gifted |
| N | 16 | 16 | 19 | 11 | 22 | 22 | 18 | 14 |
| $\Sigma X$ | 602 | 1337 | 1224 | 1032 | 1227 | 1985 | 1203 | 1275 |
| M | 37.63 | 83.56 | 64.42 | 93.82 | 55.77 | 90.23 | 66.83 | 91.07 |
| $\Sigma X^2$ | 29978 | 114157 | 88624 | 97994 | 74955 | 183891 | 90927 | 117031 |

## TABLE XXIII

### "RATE OF LEARNING" SCORES FOR AVERAGE AND GIFTED SUBJECTS ON THE TONAL CONFIGURATIONS

| | 4th Grade | | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|
| | Sequence A | | Sequence B | | Sequence A | | Sequence B | |
| | Average | Gifted | Average | Gifted | Average | Gifted | Average | Gifted |
| N | 16 | 16 | 18 | 9 | 22 | 17 | 18 | 11 |
| $\Sigma X$ | 97.2 | 533.4 | 262.5 | 601.1 | 299.5 | 1016.4 | 309.4 | 528.2 |
| M | 6.08 | 33.34 | 14.58 | 66.79 | 13.61 | 59.79 | 17.19 | 48.02 |
| $\Sigma X^2$ | 929.44 | 35215.4 | 12505.9 | 53773.8 | 12985 | 82015.8 | 14265.7 | 36710.6 |

| | | | |
|---|---|---|---|
| 1. | 4th grade, | average, | Sequence A |
| 2. | 4th grade, | average, | Sequence B |
| 3. | 4th grade, | gifted, | Sequence A |
| 4. | 4th grade, | gifted, | Sequence B |
| 5. | 6th grade, | average, | Sequence A |
| 6. | 6th grade, | average, | Sequence B |
| 7. | 6th grade, | gifted, | Sequence A |
| 8. | 6th grade, | gifted, | Sequence B |

t12 = 3.4689, there is a significant difference between the means at the .01 percent level.
t34 = 1.9550, there is no significant difference between the means.
t56 = 1.6422, there is no significant difference between the means.
t78 = .1897, there is no significant difference between the means.

2. Gifted pupils in both grades already appear to understand these concepts and derive comparatively little benefit from practice since their scores are already relatively high.
3. Average sixth-grade pupils do not show significant improvement as a result of practice, but this may be due to a number of factors such as a general lack of interest in reading music; not fully understanding the concepts inolved in music reading; and lack of experience in reading independently. This study does not attempt to identify the factors which may be present.

The writer was particularily interested in attempting to identify the rate at which pupils in these two groups were able to learn the five test items. For this purpose a "rate-of-learning" score, described earlier, was used as a means of identification. This score is based on the gains made by the subject, these gains being divided by the number of trials taken by the subject as an indication of the rate at which the gains were made. The initial trial was not included in the total number of trials. This score did have several disadvantages, but the writer was not able to find a scoring system which could be adapted to the present study. The disadvantages are as follows:

1. The subject earning a perfect score on the initial trial cannot improve his score and consequently receives an RL score of 0. In treating the data, the writer arbitrarily excluded those subjects for whom no rate of learning would be possible.
2. The subject who either made no gains from the initial trial to the final trial or who earned a lower score on the final trial than on the initial trial, received an RL score of 0. These subjects are included in the data because the opportunity for learning was always present. The reason for assigning a 0 to the subject whose score became poorer was a device to avoid the problems inherent in working with negative values.
3. The RL score is based on the total test and is not, as is the case with the mean "percent correct" score, an average of the performance on each item. This works to the disadvantage of the subject who requires many trials to learn a single item, the other items being correct within the first few trials. In view of the totality of the task which involved learning the song, the writer decided to score the tonal configurations as a group rather than individually. One reason for the difficulty encountered in scoring the tasks in this study is due to the complex nature of the tasks themselves. Research activity in the field of music although calling attention to the existence of a scoring problem, fails to recommend more adequate methods of scoring.

The highest RL score a subject could earn is 100, possible only if the subject needed only a second trial on the test to learn the items not responded to correctly on the initial trial. There were 9 fourth-grade subjects and 12 sixth-grade subjects with RL scores of 100. An additional 3 fourth-grade subjects and 8 sixth-grade subjects did not have an RL score because they made correct responses to all the items on the first trial of the test. Table XXIII summarized the RL scores for the sample. The difference between the two groups in each grade is quite evident, with the mean for gifted subjects being four to five times as large as that for average subjects. The table also shows that the means for subjects in Sequence B were higher than the means for the parallel group in Sequence A, with the exception of the gifted pupils in sixth grade.

To ascertain the degree to which practice on a related task influenced the RL scores of subjects in Sequence B, "t" tests for the significance of the differences in the means were made. The "t" tests are shown at the top of the next page.

These data parallel the results of the "t" tests made to test the significance of the differences between the means of the "percent correct" scores, a result which was predictable because of the similarity of the two kinds of scores.

The reader will recall that for Phase I the differences in mean scores for fourth and sixth grades had given a t of 1.8397 and it was concluded that the differences were not significant. The 138 subjects in the present sample were considered only in terms of grade; no differentiation being made between average and gifted or between the two sequences. The t test gave a t of .9169

| | | | |
|---|---|---|---|
| 1. | 4th grade, | average, | Sequence A |
| 2. | 4th grade, | average, | Sequence B |
| 3. | 4th grade, | gifted, | Sequence A |
| 4. | 4th grade, | gifted, | Sequence B |
| 5. | 6th grade, | average, | Sequence A |
| 6. | 6th grade, | average, | Sequence B |
| 7. | 6th grade, | gifted, | Sequence A |
| 8. | 6th grade, | gifted, | Sequence B |

t12 = 1.8036, there is no significant difference between the means.
t34 = 2.1846, there is a significant difference between the means at the .05 percent level.
t56 = .5356, there is no significant difference between the means.
t78 = .8592, there is no significant difference between the means.

These data parallel the results of the "t" tests made to test the significance of the differences between the means of the "percent correct" scores, a result which was predictable because of the similarity of the two kinds of scores.

The reader will recall that for Phase I the differences in mean scores for fourth and sixth grades had given a t of 1.8397, and it was concluded that the differences were not significant. The 138 subjects in the present sample were considered only in terms of grade; no differentiation being made between average and gifted or between the two sequences. The t test gave a t of .9169 and it was again concluded that the difference between the two grades, in terms of mean RL scores, was not significant. The mean for the fourth grade was 26.17; for the sixth grade, 31.67.

It will be remembered that in Phase I the scores for the three schools showed a significant difference existing between School B and the other two schools, but that the difference between Schools A and C was not significant. The sample for Phase I was a random sample while the subjects for Phase II represented two distinct levels of competence, with only the average group selected on a random basis. To ascertain whether School B continued to rank the lowest in terms of achievement, the RL scores on the tonal configuations for the three schools, combining grades and groups, were analyzed. Analysis of variance showed that there was a significant difference between the three schools at the 5% level, F = 4.5328. This data is presented in Table XXIV. The "t" tests for the significance of the differences between schools showed the following:

1. Between Schools A and B, t = 1.3265, and the difference is not significant.
2. Between Schools A and C, t = 2.9211, and the difference is significant at the one percent level.
3. Between Schools B and C, t = 1.764, and the difference is not significant.

The means of these three schools were: School A, 21.47; School B, 26.99; and School C, 42.09. The data tends to support the following conclusions:

1. Subjects in School C performed at a high level of competence in both Phases I and II.
2. Subjects in School A performed similarily to those in School C for Phase I but on Phase II their performance more closely paralleled that of School B.
3. Subjects in School B ranked lowest in performance on Phase I but paralleled School A on Phase II.
4. The change in position of School A for the two phases may be due to the differences in samples. It is also possible that Phase II represents a more accurate evaluation of the pupils in the schools than did Phase I. If this is the case, then it may be concluded that subjects in Schools A and B are similar while the subjects in School C are different, in terms of general competence in reading music.

2. Test Results: Song—For Sequence A, the song followed a 15-minute learning period of the five tonal configurations, for Sequence B it was the initial task. Analysis of the "percent correct" scores, with 100 the highest possible score, showed a total of only 10 perfect scores. It will be remembered that in scoring the song, the score unit was the tonal configuration within the song. Two gifted fourth-grade subjects, one in each sequence, and 8 gifted sixth-grade subjects in Sequence A account for the perfect scores. This is a partial indication that learning a song is a more difficult task than learning individual tonal configurations, for which 32 perfect scores were obtained. The data in Table XXV, again, indicate the superiority of the gifted subjects. Comparisons between the groups for the two sequences show the following:

1. For the 4th grade, average, Sequence A with Sequence B, t = 3.0018, indicating that the difference between the means is significant at the one percent level.
2. For the 4th grade, gifted, Sequence A with B, t = 2.0170, indicating no significant difference between the means.
3. For the 6th grade, average, Sequence A with B, t = 5.2728, indicating a significant differ-

## TABLE XXIV

### ANALYSIS OF VARIANCE BETWEEN SCHOOLS FOR THE TONAL CONFIGURATIONS OF PHASE II

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square |
|---|---|---|---|
| Total | 144347.7997 | 127 | |
| Schools | 9760.9410 | 2 | 4880.4705 |
| Pupils in Schools | 134586.8587 | 125 | 1076.6949 |
| | | | (F = 4.5328) |

## TABLE XXV

### "MEAN PERCENT CORRECT" SCORES FOR AVERAGE AND GIFTED SUBJECTS ON THE SONG

| | 4th Grade | | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|
| | Sequence A | | Sequence B | | Sequence A | | Sequence B | |
| | Average | Gifted | Average | Gifted | Average | Gifted | Average | Gifted |
| N | 16 | 16 | 19 | 11 | 22 | 22 | 18 | 14 |
| $\Sigma X$ | 690 | 1283 | 484 | 738 | 1466 | 2012 | 667 | 1019 |
| M | 43.13 | 80.19 | 25.47 | 67.09 | 66.64 | 91.45 | 37.06 | 72.79 |
| $\Sigma X^2$ | 36572 | 106029 | 15432 | 53238 | 103322 | 185812 | 30909 | 77189 |

ence at the .001 level.

4. For the 6th grade, gifted, Sequence A with B, t = 4.5811, indicating a significant difference between the means at the .001 level.

These results support the following conclusions regarding the importance of practice in learning to read a song:

1. Only the gifted fourth-grade subjects did not appear to be influenced by the presence or absence of practice on tonal configurations prior to learning a song; this is, of course, true only for this particular sample.
2. For all other groups, especially at the sixth-grade level, prior practice on tonal configurations resulted in significant improvement in reading the song. This was not evident when the scores on tonal configurations were compared on the basis of prior practice on the song. This tends to indicate that the reading of songs should be preceded by some prior related activity, not necessarily learning tonal configurations, so the subject moves from the simple to the complex and is aware of what is taking place.

No attempt was made to compare the scores for the song with those of the tonal configurations, because, in view of the different nature of the two tasks, such comparisons could not be completely justified.

The "rate of learning" scores for the song were analyzed in the same manner as those for the tonal configurations. Only two gifted sixth-grade subjects who were in Sequence A had a perfect score on the initial trial of the song. Eight additional gifted subjects had an RL score of 100, as compared with the 21 subjects having RL scores 100 on the tonal configurations. Table XXVI gives the data for the eight groups, and it is again evident that gifted subjects perform at a much higher level. With the exception of the fourth-grade gifted subjects, all groups in Sequence A showed higher RL scores than those in Sequence B. When one compares Table XXVI with Table XXIII it may be seen that the RL scores on the song are considerably lower than the RL scores on the tonal configurations. The t tests between the various groups comparing the two sequences showed the following:

1. For 4th grade, average, Sequence A with B, t = 1.0677, indicating no significant difference at the 5 percent level.
2. For 4th grade, gifted, Sequence A with B, t = .1847, indicating no significant difference.
3. For 6th grade, average, Sequence A with B, t = 3.6322, indicating a significant difference at the .001 percent level.
4. For 6th grade, gifted, Sequence A with B, t = 2.6444, indicating that the differences between mean RL scores is significant at the .02 percent level.

The results agree, in general, with the t tests made for the "percent correct" scores on the song and would indicate that the more musically mature and experienced sixth-grade subject is better able to profit from certain types of related practice than the fourth-grade subject.

The four groups in each grade were combined into a single group and the RL scores for the two grades compared. The data showed a mean of 13.15 for fourth grade; 21.19 for sixth grade. The t test gave a t of 2.0532 and the difference between the means of the two grades was considered to be significant at the 5 percent level. This would also support the conclusion that the maturity of the sixth-grade subject gives him an advantage when the rate of learning is considered.

It has been established that for Phase I and for the tonal configurations of Phase II, the differences between the three schools was significant. The means of the RL scores on the song, combining all groups within each school, were as follows: School A, 15.09; School B, 17.48; and School C, 19.10. Analysis of variance showed that these differences were not significant at the 5 percent level, F = .4182.

It may be concluded that the samples from the three schools vary considerably when comparisons are made in terms of achievement scores, such scores reflecting previous musical training and experience; but that these same samples do not vary significantly when comparisons are made in terms of the ability of the subjects to learn the song. The tasks which served as a basis for comparing schools in Phase I and the tonal configurations of Phase II evidently gave the subjects a greater opportunity to utilize previous experiences than did the learning of the song. This also supports the generally accepted point of view that reading a song is a highly complex process.

3. Test Results: Identity of Tonal Configurations—One of the major purposes of this exploratory study was to test the hypothesis that tonal configurations which are readily perceived in isolation tend to retain their identity when presented in the usual context of a song. The data used for this aspect of the study is the "percent correct" score for each tonal configuration in the five-item test and the same score for each item in the song. The product-moment correlations for all groups on each item were calculated and the basis for comparison is the size of the resultant "r." The raw data will not be used in this section of the report in order to simplify and condense the presentation. For the same reason, no attempt is made to retain the identity of the individual schools.

Preliminary correlations for the complete test were calculated, with the "percent correct" score for the five-item test compared to the "percent correct" scores for the tonal configurations with-

## TABLE XXVI

## "RATE OF LEARNING" SCORES FOR AVERAGE AND GIFTED SUBJECTS ON THE SONG

| | 4th Grade | | | | 6th Grade | | | |
|---|---|---|---|---|---|---|---|---|
| | Sequence A | | Sequence B | | Sequence A | | Sequence B | |
| | Average | Gifted | Average | Gifted | Average | Gifted | Average | Gifted |
| N | 16 | 16 | 19 | 11 | 22 | 20 | 18 | 14 |
| $\Sigma X$ | 91.7 | 314.4 | 81.3 | 236.1 | 277.4 | 966.0 | 99.3 | 302.6 |
| M | 5.73 | 19.65 | 4.28 | 21.46 | 12.61 | 48.3 | 5.52 | 21.61 |
| $\Sigma X^2$ | 836.99 | 14006 | 565.5 | 12892.8 | 4520.9 | 70862.7 | 957.55 | 9192.1 |

## TABLE XXVII

## ANALYSIS OF VARIANCE BETWEEN SCHOOLS FOR THE SONG OF PHASE II

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square |
|---|---|---|---|
| Total | 65770.5199 | 135 | |
| Schools | 411.0686 | 2 | 205.5343 |
| Pupils in Schools | 65359.4513 | 133 | 491.4244 |
| | | | (F = .4182) |

in the song. This procedure gave the following correlations for the combined schools:

| | | | | |
|---|---|---|---|---|
| 1. | 4th, | average, | Sequence A | r = .32 |
| 2. | 4th, | average, | Sequence B | r = .71 |
| 3. | 4th, | gifted, | Sequence A | r = .42 |
| 4. | 4th, | gifted, | Sequence B | r = .44 |
| 5. | 6th, | average, | Sequence A | r = .58 |
| 6. | 6th, | average, | Sequence B | r = .37 |
| 7. | 6th, | gifted, | Sequence A | r = .41 |
| 8. | 6th, | gifted, | Sequence B | r = .60 |
| 9. | 4th, | average and gifted, | Sequence A | r = .72 |
| 10. | 4th, | average and gifted, | Sequence B | r = .74 |
| 11. | 6th, | average and gifted, | Sequence A | r = .76 |
| 12. | 6th, | average and gifted, | Sequence B | r = .61 |

It will be noted that the correlations for the fourth-grade average and the sixth-grade gifted are considerably higher for groups in Sequence B than for those same groups in Sequence A. The reverse is true for the sixth-grade average groups, and there is little difference between the two sequences when the fourth-grade gifted groups are considered. When the two ability groups are combined, the differences between sequence for the fourth grade become less evident and the sixth grade shows a higher correlation for Sequence A. It is interesting to note that these correlations are higher than those obtained for the individual ability groups. For Sequence A, fourth-grade average was .32 and fourth-grade gifted was .42, but when the two groups are combined the "r" becomes a .72. It is the opinion of the writer that these higher correlations result from the fact that the average subjects are at the lower end of the distribution and the gifted subjects at the upper end, thus yielding a higher correlation than when each group is treated separately.

These preliminary correlations show that there is a tendency for tonal configurations to retain their identity. In order to present more complete data, the same kinds of correlations were calculated for each of the five tonal configurations used in Phase II. These will be discussed individually and then summarized at the conclusion of the section. In preference to a summary table, the means for each item, together with the correlations, will be presented at the time each item is considered. Mx will designate the mean of the item when presented in the five-item test, and My will refer to the mean of the same item as it appeared in the song (see top of next column).

Item 1: $(\underline{S}\overline{D}\overline{M})$

Sequence A

| | Average Mx | Average My | Gifted Mx | Gifted My |
|---|---|---|---|---|
| 4th | 23 | 66 | 71 | 89 |
| | r = .10 | | r = .08 | |
| 6th | 49 | 87 | 78 | 99 |
| | r = .45 | | r = -.16 | |

Sequence B

| | Average Mx | Average My | Gifted Mx | Gifted My |
|---|---|---|---|---|
| 4th | 74 | 64 | 94 | 87 |
| | r = -.56 | | r = .44 | |
| 6th | 72 | 61 | 88 | 89 |
| | r = .37 | | r = -.04 | |

This item was the first measure of the song and one would anticipate, because of its unique character, that subjects would experience little difficulty in recognition. The means of the various groups show that sixth-grade subjects are more familiar with the item than are fourth-grade subjects, the exception being the average sixth-grade group in Sequence B. There is considerable variation for the correlations and only three of these appear to show that subjects recognized the item when it appeared in a context different from its initial presentation. For Sequence A the average sixth-grade group seemed to be most aware of the item; for Sequence B the average sixth-grade and gifted fourth-grade groups are consistent. The high negative correlation for the average fourth-grade group, Sequence B, indicates considerable shifting of position from task to task. This shifting is not unusual but it does present problems of interpretation. The remaining four correlations are so low as to imply that little identification took place.

Item 2 is the simplest item of the five and occurred twice within the song.(next page). Treatment I is the comparison of the individual item (Mx) with the first appearance of the same item in the song (My). Treatment II compares the individual item with the second appearance of the item in the song (Mz), and Treatment III shows the relationship of the two items (My and Mz) as they appear within the song.

All groups showed relatively high means on the item, the lowest means are on the second appearance of the item in the song for the average sixth-grade group in Sequence B. In general, the individual item shows higher means than the item within the song. For Treatment I, the only group to show consistent recognition is the gifted fourth grade, Sequence A. The data showed that the first appearance of the item in the song tended to conceal its true identity. The means for the second

Item 2: (MRD)

| | Sequence A | | | | | | Sequence B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Average | | | Gifted | | | Average | | | Gifted | | |
| **Treatment I** | Mx | My | r | Mx | My | r | Mx | My | r | Mx | My | r |
| 4th | 62 | 64 | .10 | 91 | 98 | .93 | 76 | 44 | .02 | 83 | 76 | .28 |
| 6th | 69 | 85 | .20 | 96 | 98 | .29 | 75 | 53 | -.03 | 92 | 92 | -.14 |
| **Treatment II** | Mx | Mz | r | Mx | Mz | r | Mx | Mz | r | Mx | Mz | r |
| 4th | 62 | 33 | -.17 | 91 | 76 | .38 | 76 | 15 | .41 | 83 | 67 | .53 |
| 6th | 69 | 58 | .32 | 96 | 90 | .08 | 75 | 19 | .34 | 92 | 67 | .02 |
| **Treatment III** | My | Mz | r | My | Mz | r | My | Mz | r | My | Mz | r |
| 4th | 64 | 33 | .79 | 98 | 76 | .52 | 44 | 15 | .48 | 82 | 67 | .93 |
| 6th | 85 | 58 | .22 | 98 | 90 | .28 | 53 | 19 | .30 | 92 | 67 | .26 |

appearance in the song, Treatment II, are uniformly lower for each group than the means for the first appearance. There is a slightly higher amount of recognition for the second appearance of the item as evidenced by the five "r's" above .30. Treatment III shows that two appearances of the same item within a song tend to increase the degree of recognition which takes place.

All "r's" are positive, and five are .30 or higher. No single group shows a consistent performance on the item, the higher "r's" shifting from group to group for each treatment. The lack of stability suggests that subjects are not aware that the item exists as a note-grouping, but are reading it as unrelated individual notes. The two appearances within the song emphasize the grouping because the item has the same rhythmic structure in both instances and is a complete measure, the measure bars setting it apart as an entity.

The data also show that the "r's" in Treatments I and III are higher for Sequence A than for Sequence B; while in Treatment II, Sequence B is the preferred sequence. This suggests that for this item those subjects having prior practice on the individual items carried over more recognition than the subjects who began with the song and related this experience to the individual items. The data also show that for Sequence B the fourth grade is more apt to recognize the item in a different context than is the sixth grade. The additional experience may cause the sixth-grade subject to read less carefully and "invent" a response which seems more appropriate, in terms of the total situation, than the item actually presented to him.

Item 3: (SMF̄R)

| | Average | | | Gifted | | |
|---|---|---|---|---|---|---|
| **Sequence A** | Mx | My | r | Mx | My | r |
| 4th | 28 | 37 | .44 | 75 | 78 | .40 |
| 6th | 45 | 61 | .20 | 93 | 95 | -.16 |
| **Sequence B** | Mx | My | r | Mx | My | r |
| 4th | 40 | 8 | .70 | 80 | 42 | .64 |
| 6th | 58 | 23 | .39 | 92 | 64 | .73 |

Although only the gifted fourth grade in Sequence A shows high mean scores on the item, all but one of the "r's" is positive, with six being .39 or higher. The disjunct, but sequential characteristic of this item appears to simplify the recognition process since subjects remember the uniqueness of the item even though they may not always be able to sing it correctly. The fourth-grade subjects show a much higher degree of recognition and greater consistency than do the sixth-grade subjects, although the latter earned higher means throughout.

For this item the subjects in Sequence B show a higher degree of recognition than subjects in Sequence A which suggests that the appearance within a song is more meaningful than an isolated presentation.

Item 4: (F$\underline{\text{M}}\underline{\text{R}}\underline{\text{D}}$)

Sequence A

| | Average | | | Gifted | | |
|---|---|---|---|---|---|---|
| | Mx | My | r | Mx | My | r |
| 4th | 51 | 22 | .25 | 91 | 54 | .18 |
| 6th | 64 | 44 | .17 | 90 | 75 | .65 |

Sequence B

| | Average | | | Gifted | | |
|---|---|---|---|---|---|---|
| | Mx | My | r | Mx | My | r |
| 4th | 78 | 10 | .35 | 97 | 47 | .14 |
| 6th | 56 | 24 | .15 | 93 | 52 | .30 |

Item 4 proved to be more difficult in the song than as an individual item and this is shown by the lower My values. The chief reason for this was because it began the interval of a fourth above the final tone of the previous measure, and subjects experienced difficulty in locating the correct tone. Although only three of the correlations are .30 or higher, all "r's" are positive, indicating recognition of the item by many subjects. The size of the "r's" for the various groups does not show a consistent pattern, nor does one sequence seem to be preferred to the other although the gifted subjects in Sequence A do better than their peers in Sequence B.

The means of the gifted subjects in both sequences continues to exceed the means of the average subjects, a behavior that has been consistent throughout the testing (Item 5). The identity of the item appears to be retained since all "r's" are positive. For Sequence A the fourth-grade subjects, with higher "r's" than sixth-grade subjects, apparently are more aware of the individual character of the item even though their mean scores are lower. In Sequence B, the sixth grade has the higher "r's". The data show that subjects within most groups shift their position from task to task, a shifting that has been evident in the analysis of each item.

Item 5: (S$\bar{\text{L}}\underline{\text{S}}\underline{\text{F}}\underline{\text{M}}$)

Sequence A

| | Average | | | Gifted | | |
|---|---|---|---|---|---|---|
| | Mx | My | r | Mx | My | r |
| 4th | 19 | 31 | .50 | 84 | 86 | .34 |
| 6th | 51 | 64 | .39 | 95 | 97 | .13 |

Sequence B

| | Average | | | Gifted | | |
|---|---|---|---|---|---|---|
| | Mx | My | r | Mx | My | r |
| 4th | 54 | 14 | .44 | 100 | 70 | 0 |
| 6th | 73 | 42 | .49 | 91 | 81 | .72 |

4. Summary of The Findings—1. The average fourth-grade and gifted sixth-grade subjects in Sequence B show a higher degree of recognition of tonal configurations than do these same groups in Sequence A when the correlations between items on the song and individual items is considered. This suggests that these subjects identify tonal configurations presented individually more readily if they have had previous opportunity to observe those configurations in a song. It also implies that for these same groups, practice on tonal configurations does not lead to greater recognition of these same configurations in the song. The reverse holds for the average sixth-grade subjects in the sample.

2. The subjects within all groups do not perform consistently, as evidenced by the variation in the size of the "r's" for each item. The factors which cause this shift cannot be identified at the present time, but it is reasonable to assume, in view of the reliability of the test, that the effort and attitude of the subject are significant factors.

3. The data from the present study does not offer strong support for the point of view that tonal configurations tend to retain their identity in different contexts unless one is willing to accept

the coefficient of correlation obtained for the combined gifted and average subjects in each sequence as sufficient evidence of such retention. These were all relatively high, .61 to .76, and may represent a more dependable value than the "r's" obtained for each of the sub-groups.

4. The item, if it is to retain its identity from one situation to another, must first be learned. The testing carried on in Phase II did not result in such learning for the majority of the subjects. Shorter, more numerous practice sessions would eventually result in learning the items and a second task presented at that time might have significantly raised the "r's" for the groups.

5. The variability of the "r's" makes it impossible to identify any one sequence of tasks as the most desirable for the development of item recognition. The addition of a third task might well have provided the kinds of data needed to make such a judgment.

6. For all items, the gifted subjects in each grade for each sequence had higher scores than did the average subjects. In 20 of the 28 comparisons of scores the sixth-grade gifted subjects scored higher than the fourth-grade gifted. This is not the same when the "size" of the "r's" are compared. For these same groups the data show that in Sequence A the fourth-grade gifted group had higher "r's" than the sixth-grade gifted group in five out of seven such comparisons; for Sequence B the fourth-grade gifted group was higher than the sixth-grade gifted group in four out of seven comparisons.

7. The same results appear when the average groups in each grade are compared. The sixth-grade scores are higher in each case than fourth-grade scores, while the "r's" for fourth-grade groups are higher in nine of the fourteen comparisons than the "r's" of the sixth-grade groups.

8. When each item is considered individually no single item shows a uniformly high degree of recognition by all the various groups. Item 3 and the two appearances of Item 2 within the song appear to retain this identity better than any other item. The data show that for these items the fourth-grade "r's" are higher than sixth grade in 13 out of 16 comparisons. For Items 1, 4 and 5 the sixth-grade "r's" are higher than those of the fourth grade in 8 of 12 comparisons. This is a more reliable indication of what is taking place and shows that fourth-grade subjects tend to retain items that are either very simple or which have distinctive characteristics. When the item is neither of these, the sixth-grade subjects are better able to detect the subtle differences and identify the item in a new context.

9. It may be concluded that the sixth-grade subjects in Phase II of the study not only demonstrated a higher level of achievement but also showed a greater degree of item recognition for the test as a whole.

5. The Influence of Instrumental Instruction on Test Scores—It has been mentioned earlier that many music educators feel that experience in playing a musical instrument results in a higher level of music reading ability, because instrumental performance tends to clarify the abstractions of musical notation. A total of 53 subjects in Phase II had reported instrumental training of at least 12 months or more. The "mean percent correct" scores on the tonal configurations were used as the basis of comparing the pupils within groups who had training with those having no such training. The results are summarized as shown at top of next page.

The "t" tests made show that differences between the means were not considered significant at the 5 percent level and it may be concluded that the instrumental training had little effect upon the test scores of the subjects in the sample.

There are several factors which need to be considered before any conclusions can be reached:

1. The groups compared are all relatively small.
2. Piano was the instruemtn mentioned most frequently, and this may not influence music reading ability, in the sense of this study, as much as a single-line instrument like the violin.
3. The nature of the instruction may have given more emphasis to technical performance than to developing skill in reading music.
4. The instrument may not have been studied long enough to enable the child to develop an adequate concept of the process of music reading.
5. The tests may not really give a true measure of music reading ability, especially since the use of a vocal response may have served as a deterent.

For these reasons the writer is not willing to go beyond the limitations of the present study.

6. Relationship Between Scores on Tonal Configurations and the Teacher Ratings and Kwalwasser Tests—The writer has already discussed the problem associated with the identification of the musically gifted children. There are many possible definitions of gifted children and this study defined the musically gifted subject as the child whose musical achievement was superior to that demonstrated by the majority of his classmates. This superior achievement results from favorable attitudes toward musical activities, general knowledge about music, and performing skill in singing and playing instruments. The experienced teacher, having contact with the children over an extended period of time, should be able to evaluate those outstanding children in each group. The Kwalwasser Music Test does not measure achievement but seeks to identify native capacity.

| Group | Without Training N | Without Training Mean | With Training N | With Training Mean | "t" |
|---|---|---|---|---|---|
| 4th Average, Sequence A | 15 | 39.3 | 1 | 13 | |
| 4th Average, Sequence B | 16 | 66.8 | 3 | 85.3 | 1.3536 |
| 4th Gifted, Sequence A | 4 | 79.3 | 12 | 85.0 | .7763 |
| 4th Gifted, Sequence B | 6 | 90.3 | 5 | 98.0 | 1.1936 |
| 6th Average, Sequence A | 17 | 54 | 5 | 57.6 | .3339 |
| 6th Average, Sequence B | 12 | 66 | 6 | 68.5 | .1952 |
| 6th Gifted, Sequence A | 6 | 84.0 | 16 | 92.5 | 1.1961 |
| 6th Gifted, Sequence B | 5 | 91.8 | 9 | 90.7 | .2330 |

Data presented earlier in the report showed the correlations between the teacher's ratings and the Kwalwasser Test to be relatively low for all schools. This is to be expected because the two criteria do not measure exactly the same things. To ascertain which criterion measure, if any, had the closest relationship to the reading of tonal configurations in Phase II, the product-moment correlations were computed for the average and gifted subjects in each school. These correlations are represented as AB and AC:

1. A represents the scores on tonal configurations.
2. B represents the scores on teacher ratings.
3. C represents the scores on the Kwalwasser test.

These "r's" are given for each grade in each school as follows:

| School | Grade | Av. or Gifted | rAB | rAC |
|---|---|---|---|---|
| 1. A | 4 | A | .60 | -.06 |
| 2. A | 4 | G | .21 | -.10 |
| 3. A | 6 | A | .08 | -.14 |
| 4. A | 6 | G | .48 | -.36 |
| 5. B | 4 | A | .79 | .17 |
| 6. B | 4 | G | -.03 | .59 |
| 7. B | 6 | A | -.28 | -.07 |
| 8. B | 6 | G | -.29 | .17 |
| 9. C | 4 | A | -.25 | .42 |
| 10. C | 4 | G | .03 | .17 |
| 11. C | 6 | A | .13 | -.69 |
| 12. C | 6 | G | -.10 | -.09 |

These data show a lack of consistency, not only from school to school but for the various groups as well. The data show the following:

1. Neither of the criterion measures are able to show a consistent relationship with the task used in Phase II, although the teacher's rating correlates higher than the Kwalwasser for all groups in School A.
2. For School B the preference is given to the Kwalwasser test although the correlations are so low as to indicate that such a basis is not reliable. School C is similar to School B.
3. When the gifted subjects are considered, the higher "r's" occur for the Kwalwasser in four out of six such comparisons.
4. When the average subjects are considered, the higher "r's" occur for the teacher's ratings.

On the basis of the data obtained, the writer finds himself in agreement with the many other investigators who have attempted to establish some basis for identifying musically gifted individuals, that this basis is yet to be discovered. It is possible that the most adequate basis for making such identification is the performance on the tasks in Phase II rather than outside criteria. It is also quite possible that the absence of a large number of significantly high "r's" and the lack of a consistent pattern shows that teachers are not in a position to adequately evaluate the music reading ability of their pupils because there is comparatively little opportunity for individual attention within the classroom situation.

## V. Summary and Conclusions

### Summary of Major Findings Based on the Analysis of Data

1. There is no significant difference between boys and girls in terms of the ability to read music.

2. For Phase I there were no significant differences between the scores earned by fourth-, fifth-, and sixth-grade subjects. For Phase II there were no significant differences between fourth and sixth grades either in terms of general scores or in terms of the rate of learning scores for tonal configurations. However, the RL scores on the song showed a significant difference between these two grades.

3. When "percent correct" scores were used, there were significant differences between the three schools; School C, while comparable to School A in Phase I, was higher than the other

two schools in Phase II. School B ranked lower on Phase I and was comparable to School A on Phase II. There were no differences when RL scores on the song were used as a basis for comparison.

4. The analysis of data for Phase I showed:
   a. The subjects experienced considerable difficulty in reading tonal configurations which commonly appear in the songs they sing.
   b. The common reading errors revealed a general tendency to respond to the outward shape of the configuration, but to but to substitute a similar, more familiar one being read.
   c. Considerable guessing takes place when subjects attempted to read, as evidenced by the number of different errors which were made.
   d. The subjects performed at a much higher level on the aural part of the test than on the visual part. There were few common errors, the major cause for error arising out of the subjects' apparent inability to retain the sound of the starting pitch throughout the presentation of the item.
   e. The relationship between the responses to the aural and visual presentations of an item was not as high as expected, "r's" of .50 to .70. This is, however, not as accurate a comparison as might have been made.

5. The teacher's ratings of subjects were compared with the results of the Kwalwasser test and showed very low correlations.

6. Gifted subjects in Phase II consistently did better than average subjects on each task. They knew more about reading music, as evidenced by the "percent correct" scores, and learned the material at a rate three to four times faster than average subjects.

7. The song was a more complex task than the individual tonal configurations, regardless of the sequence of tasks.

8. Comparisons between sequences for the various groups, using the scores earned on tonal configurations as a basis for comparison, showed the following:
   a. The "percent correct" scores for groups in Sequence B, with prior practice on the song, were not significantly higher as a result of this practice than the scores for groups in Sequence A who did not have practice. Only the fourth-grade average group seemed to benefit from
   b. The RL scores for groups in Sequence B were not significantly higher than the scores for groups in Sequence A except for the gifted fourth-grade subjects.

9. Comparisons between sequences for the various groups, using the scores earned on the song showed the following:
   a. The "percent correct" scores for groups in Sequence A, with prior practice on tonal configurations, were significantly higher as a result of this practice than the scores for groups in Sequence B who had no practice. Only the gifted fourth-grade group seemed not to benefit from such practice.
   b. The RL scores for sixth-grade groups in Sequence A were significantly higher than the scores for the sixth-grade groups in Sequence B. There were, however, no significant differences between fourth-grade groups in the two sequences.

10. Analysis of the degree to which tonal configurations retain their identity from one situation to another showed that:
   a. Average fourth-grade and gifted sixth-grade subjects show greater recognition when the sequence from song to configuration is followed.
   b. Average sixth-grade subjects show greater recognition when Sequence A is followed.
   c. Considerable shifting of the position of the subjects from one task to another takes place within all groups. This variability makes it difficult to accurately interpret the "r's".
   d. Items 2 and 3 show fairly consistent recognition by fourth-grade groups; Items 1, 4 and 5 are more readily recognized by sixth-grade groups.

11. The scores of subjects in this study who reported instrumental training were not significantly higher than the scores of subjects who did not have such training.

12. The identification of gifted and average subjects on the basis of teacher's ratings and the Kwalwasser test does not result in a high level of reliability in terms of this study.

## Implications of the Study

1. The time allowed for the tasks in Phase II was not sufficient to result in complete learning by most of the subjects. This suggests that more learning sessions, rather than longer ones, would result in greater understanding of the concepts of musical notation.

2. Accurate reading of musical notation is more likely to result when greater emphasis is placed upon understanding the significance of the notation rather than upon mere "imitation" of musical materials.

3. The musically superior child needs to be given more opportunities to develop music reading competence. The data show both fourth- and sixth-grade superior children to learn at a much

more rapid rate than average children, a fact that must be considered in planning the music activities for the grades.

4. The data did not clearly establish that one sequence of tasks is superior to the other in developing music reading competence. It did show, however, that although prior practice on the song did not appear to have much effect upon the scores obtained for learning individual configurations, learning the song was made more effective through prior practice on tonal configurations. The fact that the song is usually approached as a complete entity, as it should be, makes it difficult for children to then recognize the individual elements of the song when they are isolated.

5. If the music reading is approached by means of the configuration, greater emphasis needs to be placed upon helping children recognize the shape and design of such configurations. Subjects in this study were unable to identify the configurations in the song which they had previously learned as individual items until they were asked to make such an identification.

6. The results of the study showed that these sixth-grade children did not perform significantly better than the fourth-grade children, although they clearly demonstrated a faster learning rate for the song. This implies that very little actual growth in music reading skill takes place beyond the fourth-grade level, because the children are not given the necessary musical tools with which to work. The tools in this sense of the work are the knowledges and understandings about musical notation which are developed through independence of action.

7. Instrumental instruction did not seem to be a factor in the music reading abilities of either the average or gifted children in this study. The training will exert a greater influence if greater emphasis is given to the meaning of the musical notation rather than to mere mechanical responses to visual stimuli. It is highly probable that instrumental training at this level is more concerned with developing control over the instrument so that he is not able to give his attention to other problems of performance.

8. The procedures followed in this study are adequate, but improvements need to be made in methods of scoring and evaluating the data.

9. The problem of the visual perception of music symbols needs to be extended to ascertaining whether tonal configurations learned in one key are readily perceived when presented in other tonalities where the notation is different.

10. The relationship between auditory and visual perception needs to be more clearly established. Music educators recognize the interdependence of the two, but the degree to which one depends upon the other needs to be identified. Many reading errors made in this study were caused by the subject's inability to "hear internally" what the stimulus should sound like and, consequently, were unable to control the voice.

## REFERENCES

1. Bean, Kenneth, "An Experimental Approach to the Reading of Music," Psychological Monographs, Vol. L, No. 6 (1938).
2. Christ, W. B. Reading of Rhythm Notation Approached Experimentally According to the Techniques and Principles of Word Reading, unpublished Ph. D. dissertation, Indiana University, 1953.
3. Flanagan, John C. "Considerations in the Selection of Test Items and a Short Method of Estimating the Product-Moment Coefficient from Data at the Tails of the Distribution," Journal of Educational Psychology, XXX (1939), pp. 674-80.
4. Jacobsen, O. Irving, "An Analytical Study of Eye-Movements in Reading Vocal and Instrumental Music," Journal of Musicology, III pp. 1-32; 69-100; 133-64; 197-221; 223-26.
5. Ortmann, Otto, "On the Melodic Relativity of Tones," Psychological Monographs, Vol. XXXV, No. 1 (1926).
6. Ortmann, Otto, "Some Tonal Determinants of Melodic Memory," Journal of Educational Psychology, XXIV (1933), pp. 454-67.
7. Ortmann, Otto, "Span of Vision in Note Reading," Journal of Proceedings of the Thirtieth Annual Meeting of the Music Educators National Conference (1937), pp. 88-93.
8. Smiley, Edra M. A Study of the Musical Configurations, Symbols, Terms and Words Found in Basic Music Texts at the 4th Grade Level, unpublished Ph. D. dissertation, Indiana University, 1954.
9. Stokes, Charles F. An Experimental Study of Tachistoscopic Training in Reading Music, unpublished Ph. D. dissertation, University of Cincinnati, 1943.
10. Weaver, H. E. "Studies of Ocular Behavior in Music Reading," Psychological Monographs, Vol. LV, No. 1 (1943).
11. Weaver, H. E., and Van Nuys, Kelvin, "Memory Span and Visual Pauses in Reading Rhythms and Melodies," Psychological Monographs, Vol. LV, No. 2 (1943).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 4, June 1960)

# A STUDY OF RELATIONSHIPS BETWEEN IMPORTANCE OF A SCHOOL SUPERINTENDENT'S FUNCTIONS AND HIS PRACTICES, BASED ON AUTHORITATIVE OPINION

RAYMOND J. YOUNG
University of Michigan
BERYL DILLMAN
University of California
Santa Barbara, California

ONE OF THE most important and influential public officials in America is the public school superintendent of schools. Within the legal structure of school government extant in each state, the superintendent of schools is an executive officer directly responsible to a local board of control for the maintenance, operation, and administration of the public schools established within the district over which the controlling board has legal jurisdiction.

Early New England public schools were controlled and maintained by decisions reached in town meetings. Later as the population increased and school enrollment grew, appointed and elected representative boards of control came into being whose function it was to maintain and operate the respective local public schools. After nearly 200 years, the increased size of school enrollments had operated as one factor to precipitate increasingly complex and numerous problems related to the administration of public schools. It became difficult for members of a board of control to administer the schools as had previously been possible under frontier conditions. During the 121 years since the first city superintendent of schools was appointed at Buffalo, New York, the functions and duties of the office of superintendent of schools have evolved gradually to their present day status. Functions of the office became more numerous and complex as increasingly better and more highly trained persons were available, demanded, and employed for that position. The position has become more exacting and subject to varied and increased numbers of extra- and quasi-legal, political, economic, religious, and sociological pressures. The effects of these pressures and demands on the time of the superintendent of schools were suggested by Norton[1]* when he stated that "The school superintendency is becoming an untenable position in some large cities, and in many middle-sized and smaller communities it is not a bed of roses." The nature of the public school superintendent's functions, duties, and responsibilities is constantly evolving.

*The Evolving School Superintendents' Task and Function*—Perusal of the literature relating to the school superintendents' task seems to warrant certain conclusions pertaining to changes in emphasis and nature of concepts held regarding the office. More recently there seems to be increased emphases upon the welfare of the child as the criterion for justifying and evaluating the effectiveness of administration. Increased emphases on the responsibility and role of administration for the improvement of instruction, including attention to the selection of a well qualified staff, have been noted. It seems also, that a major evolving change of emphases in concept relates to the increased trend toward democratic policy making and problem solving simultaneously with need to delegate duties to others. This change involves wider utilization of staff, pupils, and school patrons in the community in formulating policy and solving problems. Increased emphasis is also indicated on the function of public relations.

In 1923 the Department of Superintendence of the National Education Association queried school superintendents on their perception of the importance of various functions deemed to be those of a superintendent of schools, and a similar study was repeated in 1933 by the American Association of School Administrators.[2] Data obtained in each of these two years reveal the changing concept of school superintendents relative to the importance of various functions. During that decade increased emphasis was noted on personnel problems and supervision of instruction, while a decrease in emphasis on importance of functions relating to budget, purchasing problems, census taking, and buildings and grounds was evidenced.

---

*All footnotes will be found at end of article.

In 1950, 1586 school superintendents responded to an inquiry conducted by the American Association of School Administrators concerning the actual and ideal percentage of time distribution for superintendents among eight major broad classifications of functions.[3] Data revealed that the function of "instructional leadership" was considered as the one to which superintendents do and should devote the most time.

However, school superintendents expressed need for more time emphasis on 1) instructional leadership, 2) general planning, and 3) public relations. They indicated too much time was devoted to 1) financial administration, 2) personnel administration, and 3) school plant management.

The earlier studies cited, regarding the relative importance of school superintendents' functions were based upon what school superintendents responding to a study thought. The first major effort to measure the superintendent's thinking and practice against some sort of criteria was made by Russell.[4] After synthesizing the literature related to functions of a school superintendent, he developed a list of 38 functions. This list was submitted to a jury of carefully selected experts consisting of 100 outstanding school superintendents, 60 outstanding professors of school administration and 20 other distinguished educators. Jury members were distributed throughout the United States and included all the nation's well known leaders in the area of school administration. The jury of 180 persons was asked to indicate the relative importance of the functions included on the listing.

An identical list of functions was submitted to superintendents of all (116) Oklahoma North Central Association schools. Superintendents were also asked to indicate their judgment concerning the relative importance of each function; but, in addition, they were asked to indicate the extent to which time they spent on each function was "about right," "too little," or "too much." Of the 116 superintendents included in the study, 102 or 88 percent responded. (Each of the Oklahoma superintendents was the chief administrator in districts maintaining grades 1-12.)

Responses from the jury of outstanding persons in school administration served as a body of authoritative opinion and a criterion against which the thinking and perceived practice of Oklahoma school superintendents could be measured.

Problem and Nature of This Study—The general problem of this study was to determine the relationship between the importance of administrative functions and the time expenditures of school superintendents relative to various administrative functions. Answers to several major questions were sought.

1. What is the relative importance of various administrative functions held by Illinois superintendents?
2. To what extent do the perceptions of importance regarding administrative functions held by Illinois elementary school superintendents agree with those held by experts in the field of school administration?
3. What is the relationship between the perceived time expenditures of Illinois superintendents relative to various administrative functions and their own concepts of the relative importance of various functions?
4. What is the relationship between the perceived time expenditures of Illinois superintendents relative to various administrative functions and the importance of various functions according to authoritative opinion?
5. Does the validity of perceived time expenditures of Illinois superintendents vary among schools in different categories of enrollment size?
6. How does the validity of perceived time and energy expenditures of Illinois superintendents compare with those of Oklahoma superintendents based on authoritative opinion?

In seeking answers to these questions, the study was limited to Illinois superintendents of separate elementary school districts. Superintendents of separate elementary school districts who were also serving as a superintendent of a separate high-school district were excluded. A twenty percent random sample of separate elementary districts, stratified by categories of school enrollment size, was selected for the study. Superintendents of 50 such school districts were invited to participate.

The study was in two parts. First, each of the 50 superintendents was sent a list of 38 functions and asked to indicate the importance of each function according to whether he felt it was "Extremely Important," "Significantly Important," "Moderately Important," or of "Little or No Importance." The list of functions used was identical to the one developed by Russell through a synthesis of the literature on educational administration in order that comparative data could be obtained. All 50 superintendents responded.

Secondly, after a superintendent had returned the first part of the study, he was forwarded another list of the same 38 functions and asked to indicate an evaluation of his own practices in regard to the time and energy expended relative to each function. Enough time was allowed to elapse between administration of the first and second parts of the study so that a respondent would be unlikely to recall his response to the first part. A superintendent did this by indicating for each function whether the time and energy he expended were "Too Much," "Too Little," or "About Right." All except two complete responses were received to the second part of the study. Due to the size of the sample, the manner in which the sample was selected, and the number of responses to each part of the study, one is justified in confidently generalizing from the findings to all Illinois superintendents of separate elementary school districts.

Responses to each of the two parts of the study were collated separately by category of school enrollment size. Frequencies of response on importance of functions were assigned one of four weights, and frequencies of response on amount of time and energy were assigned one of three weights. Weighted ratings were converted into weighted scores for interpretative and statistical treatment and comparative purposes. Functions were ranked in degree of importance according to magnitude of the weighted scores thus derived. The same procedure was applied to responses on perceived time and energy expenditures.

Assumptions Underlying This Study—The present study is based on the assumptions that:

1. The evolving concepts relating to the functions of the superintendent of schools and their relative importance have not changed substantially since the opinions of experts in school administration were obtained in 1954.
2. The chief difference between the administration of a small school and a large one is not in the administrator's function but rather in the generalization of duties in the former and in their specialization in the latter.
3. The functions of the public school superintendent are substantially the same regardless of school size, community economic background or social structure, and number of grades maintained by the district.
4. The primary responsibility of the school superintendent is to develop a functional educational program through the exercise of leadership as a director or boss, a chairman of planning groups, a coordinator of services, and an educational statesman. In discharging this responsibility certain functions are present which must always be taken care of, regardless of the nature of the administrative process.

Relative Importance of Administrative Functions—Data are presented in Table I showing the rank order of functions by importance according to the jury of experts, Illinois elementary school district superintendents, and Oklahoma superintendents of schools. By reading the ranks from left to right opposite a given function it is possible to determine the relative importance of a function held by each of the two groups of school superintendents compared to the jury of experts. More than casual observation is needed to determine the extent to which functions deemed to be most important by the superintendents are also those functions held to be most important by outstanding school superintendents, professors of school administration, and authors composing the jury of experts. Rank difference coefficients of correlation showing the several relationships are presented in Table II. Schools with enrollments between 200-499, 500-999, and 1000 and over were classified respectively for this study as small, medium, and large.

Table II is to be interpreted in the following manner: The coefficient of correlation between the relative importance associated with the various administrative functions by any group of superintendents listed under Column (1) and either the jury of experts (Column 3) or Oklahoma superintendents (Column 2) may be determined by reading the number appearing under either of the latter two columns and opposite the group listed under Column (1). For example, the coefficient of correlation between all Illinois superintendents included in the study and the jury of experts on relative importance of various functions is .85.

Data presented in Table II reveal that the 102 Oklahoma school superintendents responding to the study agreed more closely with authoritative opinion than did the 50 Illinois elementary school district superintendents as to the relative importance of the various functions. However, a coefficient of .90 between the relative importance of functions held by Illinois superintendents compared with that indicated by Oklahoma superintendents shows a high degree of agreement between the two groups of superintendents. The most significant finding is probably the fact that when relationships are analyzed by categories of school size it is found that the relative importance associated with various functions by superintendents tends to increasingly diverge from authoritative opinion as school enrollment size decreases. The same trend is true to a lesser extent when relationships between what Oklahoma and Illinois school superintendents think important are considered.

In general, it can be concluded that school superintendents tend to substantially agree with the experts in school administration about the relative importance of various administrative functions. Illinois superintendents of smaller separate elementary school districts do not agree as closely with authoritative opinion about the relative importance of various administrative functions as do superintendents in larger similar districts. Russell[5] also found a similar tendency for judgments on the relative importance of various functions to vary among superintendents according to size of the community in which a school was located.

Relative Importance of Functions Grouped by Area—Although the 38 functions submitted to respondents were intentionally ungrouped all functions can be classified into one of nine area groupings. Average weighted ratings were found for all functions within each area classification or group-

TABLE I

FUNCTIONS RANKED ACCORDING TO IMPORTANCE BY JURORS, ILLINOIS ELEMENTARY SUPERINTENDENTS, AND OKLAHOMA SUPERINTENDENTS

| Function of Superintendent | Rank by Jury | Rank by Illinois Elementary Superintendents | Rank by Oklahoma Superintendents |
|---|---|---|---|
| Employing of teaching, supervisory and administrative personnel | 1.5 | 1.0 | 1.0 |
| Guiding board of education in policy making | 1.5 | 3.5 | 2.5 |
| Preparing budget | 5.0 | 3.5 | 2.5 |
| Formulating school policies | 5.0 | 8.0 | 5.5 |
| Establishing and maintaining effective staff relationships | 5.0 | 2.0 | 5.5 |
| Planning long range building program | 5.0 | 18.0 | 14.0 |
| Reporting to the public | 5.0 | 16.5 | 9.5 |
| Delegating authority | 9.5 | 14.5 | 9.5 |
| Improving personal and professional qualifications of self | 9.5 | 5.0 | 5.5 |
| Providing leadership for in-service of staff | 9.5 | 14.5 | 14.0 |
| Evaluating school progress | 9.5 | 6.0 | 14.0 |
| Assigning, transferring, and dismissing personnel | 12.5 | 16.5 | 5.5 |
| Striving for democratic approach to problem solving | 12.5 | 19.5 | 14.0 |
| Initiating curriculum study and revision | 14.0 | 9.0 | 20.0 |
| Maintaining working relations with P. T. A. groups | 15.5 | 22.0 | 20.0 |
| Preparing of building plans | 15.5 | 13.0 | 9.5 |
| Supervising the improvement of instructions | 17.0 | 7.0 | 9.5 |
| Keeping operation of school in proper legal channels | 18.5 | 11.0 | 14.0 |
| Making contacts with civic, fraternal, and religious groups | 18.5 | 24.0 | 22.5 |
| Interviewing patrons and representatives of local organizations | 20.0 | 27.5 | 24.5 |
| Holding and attending workshops, clinics, institutes | 21.5 | 25.0 | 27.5 |
| Employing of clerks, custodians, other non-professional personnel | 21.5 | 19.5 | 17.5 |
| Preparing courses of study | 23.5 | 12.0 | 20.0 |
| Keeping proper records | 23.5 | 10.0 | 17.5 |
| Reporting to federal, state, and local authorities | 25.0 | 23.0 | 30.0 |
| Establishing proper rapport with public | 26.0 | 21.0 | 24.5 |
| Providing health examinations for pupils and employees | 27.0 | 31.0 | 26.0 |
| Supervising the maintenance of the school plant | 28.0 | 27.5 | 22.5 |
| Administering the transportation program | 29.5 | 32.0 | 27.5 |
| Supervising non-teaching employees | 29.5 | 30.0 | 33.0 |
| Taking of school census | 32.0 | 37.0 | 33.0 |
| Designing of record forms | 32.0 | 36.0 | 37.0 |
| Selecting library and other supplies | 32.0 | 29.0 | 30.0 |
| Purchasing general supplies (non-educational) | 34.5 | 33.5 | 33.0 |
| Supervising attendance | 34.5 | 35.0 | 30.0 |
| Administering lunch and other auxiliary programs | 36.5 | 33.5 | 35.5 |
| Dealing with pupil problems, including discipline | 36.5 | 26.0 | 35.5 |
| Teaching classes | 38.0 | 38.0 | 38.0 |

## TABLE II

## RANK DIFFERENCE COEFFICIENTS OF CORRELATION BETWEEN IMPORTANCE OF ADMINISTRATIVE FUNCTIONS BASED ON AUTHORITATIVE OPINION AND IMPORTANCE INDICATED BY SCHOOL SUPERINTENDENTS

| (1) Group of Superintendents | (2) Oklahoma Superintendents | (3) Jury of Experts |
|---|---|---|
| All Illinois Superintendents | .90 | .85 |
| Oklahoma Superintendents | 1.00 | .93 |
| Illinois Superintendents (Small Schools) | .81 | .74 |
| Illinois Superintendents (Medium Schools) | .86 | .83 |
| Illinois Superintendents (Large Schools) | .92 | .89 |

ing in order to determine which areas of an administrator's total task were deemed to be of greatest importance. Table III contains average weighted ratings by area of function for the jury of experts and both groups of superintendents. When the weighted ratings are interpreted according to the established categories it is evident that both groups of superintendents and the experts concur that functions related to the Board of Education are "Extremely Important." While the experts consider functions in the area of public relations to be "Extremely Important," both groups of superintendents indicate they consider them as only "Significantly Important." Both the experts and Illinois superintendents considered functions related to administration of personnel as "Significantly Important," while Oklahoma superintendents considered them "Extremely Important." Without including functions concerning the employment and supervision of non-teaching employees for experts and Oklahoma superintendents, the area of staff personnel administration would have an average weighted rating well above the minimum necessary to have been classified by them as "Extremely Important." One can conclude that functions relative to staff personnel are considered as one of the most important areas with which the administrator must be concerned.

All three groups concur that functions related to finance, buildings, and instruction are "Significantly Important." While the experts considered functions of a superintendent related directly to pupils and of a ministerial nature to be only "Moderately Important," both groups of superintendents rated them as "Significantly Important." Illinois superintendents and the experts agreed that functions classified as miscellaneous which included supervision of attendance, designing record forms, and administering transportation, lunch and other auxiliary programs were only "Moderately Important." Oklahoma superintendents indicated they were "Significantly Important."

The two areas of greatest discrepancy between degree of importance associated by Illinois elementary school district superintendents and the experts are those of public relations and pupil personnel administration.

Some insight into the nature of the difference between the jury of experts and Illinois superintendents on the relative importance of various areas of administrative functions can be obtained by noting the specific functions where there was a difference of five relative ranks or more between them. Below are listed the functions thought to be of greater importance by the experts than by Illinois superintendents:

Reporting to the public.
Striving for a democratic approach to problem solving.
Maintaining working relations with PTA groups.
Interviewing patrons and representatives of local organizations.
Making contacts with civic, fraternal, and religious groups.
Taking school census.
Providing leadership for in-service training of staff.
Planning long range building program.

All functions pertaining to the area of public relations appear to be considered more important by the experts. Functions accorded greater relative importance by Illinois superintendents than by the experts are listed below. The first three functions pertain to administration and improvement of the instructional program.

Initiating curriculum study and revision.
Supervising the improvement of instruction.
Preparing courses of study.
Keeping operation of school in proper legal channels.
Keeping proper records.
Establishing proper rapport with pupils.
Dealing with pupil problems, including discipline.

Whereas the experts placed a higher relative degree of importance than superintendents on "Providing leadership for in-service training of the staff," superintendents placed a higher relative importance than the experts on functions directly involving the administrator in personally bringing about curriculum improvement. The experts emphasize the leadership role of the administrator in providing and stimulating conditions related to staff improvement, whereas the administrators emphasized functions which would imply that the administrator would personally and directly bring about curricular improvement. Elementary school district superintendents placed a higher relative importance on functions relating to direct contact with pupils than did the experts.

Extent to which Superintendents Practice What They Know—Once having discovered the fact that, in general, school superintendents tend to agree with authoritative judgment about the relative importance of their various functions, it was then the task of the investigation to ascertain the extent to which the superintendents' perceived time and energy expenditures in practice coincided with what they seem to know is important. With reference to each function, superintendents indicated an evaluation of the adequacy of their own time and energy expenditures by indicating whether it was "too much," "too little," or "about right." The perceived practice has been analyzed in the light of general agreement on what is important in the task of educational administration as a superintendent of schools. An analysis of these data indicates the validity of time and energy expenditures of the

TABLE III

COMPARISON OF AVERAGE WEIGHTED RATINGS FOR EXPERTS AND SUPERINTENDENTS BY AREA OF FUNCTION

| Area of Function | Average Weighted Ratings* | | |
|---|---|---|---|
| | Jury of Experts | Illinois Superintendents | Oklahoma Superintendents |
| Board of Education | 3.6 | 3.6 | 3.6 |
| Public Relations | 3.5 | 3.0 | 3.3 |
| Personnel | 3.4 | 3.4 | 3.5 |
| Buildings | 3.3 | 3.1 | 3.4 |
| Instruction | 3.2 | 3.3 | 3.3 |
| Finance | 2.6 | 2.9 | 3.1 |
| Ministerial | 2.4 | 2.8 | 3.0 |
| Pupils | 2.0 | 2.5 | 2.7 |
| Miscellaneous | 1.9 | 2.3 | 2.7 |

*For purposes of interpretation, numerical categories of weighted ratings were established as follows:

| | |
|---|---|
| 0.0 - 1.4 | Little or No Importance |
| 1.5 - 2.4 | Moderately Important |
| 2.5 - 3.4 | Significantly Important |
| 3.5 - 4.0 | Extremely Important |

school superintendent in terms of what is important. These data are presented in Table IV which is to be interpreted as follows: The relationship between relative importance of various functions and the perceived amount of time and energy expenditures as indicated by all Illinois superintendents included in the study is -.48. This means that functions considered by these superintendents as having the greatest importance tend to have "too little" time and energy spent on them by these same superintendents and that functions of lesser importance tend to have "too much" time and energy expended on them.

Coefficients of rank difference correlation presented in Table IV indicate a significant inverse relationship exists between importance of functions according to the superintendents' and jury of experts' views about the relative importance of various functions and the time and energy expenditures relative to them. These data indicate that, as perceived by superintendents, too little time and energy on the job are devoted to the functions which both they and the experts tend to agree are of greatest importance. Too much time and energy are expended on functions of lesser importance. These facts were also found to be true of Oklahoma superintendents to a degree.

When analyses were made considering school size, it appears that this condition is more true among superintendents of schools with small enrollments than for those of medium or large size. It appears also that superintendents of schools with medium size enrollments tend to devote more time and energy to functions deemed to be of greatest importance than do superintendents in the other two enrollment size categories.

Functions Receiving "Too Little" Time and Energy—These significant findings prompted further analyses concerning the relationships between functions in regard to which superintendents indicated they were expending "too little" time and energy. Functions were ranked on which "too little" time and energy were being expended according to extent to which this was true. The ten top ranked functions on which Illinois superintendents indicated "too little" time and energy were being spent are listed below in rank order of mention:

1. Supervising the improvement of instruction.
2. Providing leadership for in-service training of staff.
3. Evaluating school progress.
4. Reporting to the public.
5. Improving personal and professional qualifications of self.
6. Preparing couses of study.
7. Initiating curriculum study and revision.
8. Holding and attending workshops, clinics, institutes.
9. Establishing proper rapport with pupils.
10. Formulating school policies.

Observation of the listing reveals that such functions are largely concerned with evaluation, curriculum development and improvement, staff and self-improvement, and relationships with the public and pupils.

Table V contains rank difference coefficients of correlation between the importance of various functions and the extent to which superintendents perceive that they devote "too little" time and energy. Data indicate that functions on which superintendents devote "too little" time and energy are those that tended to have been deemed by both experts and Illinois superintendents as most important. Again there is evidence that Illinois superintendents of medium size schools tend more than those of large and small ones to devote "too little" time and energy to a lesser extent to more important functions. The difference between the coefficients of correlation for all Illinois and Oklahoma superintendents was found to be significant at the 5 percent level of significance. This would indicate that Illinois superintendents, more than Oklahoma superintendents, perceived that they devoted "too little" time and energy to functions deemed to be of greatest importance.

Functions Receiving "Too Much" Time and Energy—Functions concerning which Illinois superintendents indicated they devote "too much" time and energy were ranked in descending order according to the frequency with which superintendents had indicated "too much." Following are listed the ten most frequently mentioned functions with reference to which superintendents perceive they devote "too much" time and energy:

1. Reporting to federal, state, and local authorities.
2. Administering lunch and other auxiliary programs.
3. Supervising the maintenance of the school plant.
4. Purchasing general supplies (non-educational).
5. Supervising non-teaching employees.
6. Keeping proper records.
7. Dealing with pupil problems, including discipline.
8. Designing of record forms.
9. Preparing of building plans.
10. Administering the transportation program.

Rank difference coefficients of correlation are presented in Table VI to indicate the relationship between the relative importance of administrative functions according to both the jury of experts and Illinois superintendents, and the functions concerning which superintendents perceive they devote "too much" time. Significant inverse relation-

TABLE IV

RANK DIFFERENCE COEFFICIENTS OF CORRELATION BETWEEN IMPORTANCE OF ADMINISTRATIVE FUNCTIONS AND TIME AND ENERGY EXPENDITURES FOR SUPERINTENDENTS

| (1) Groups Determining Importance of Functions | (2) Amount of Time and Energy Expenditures | | | |
|---|---|---|---|---|
| | All Illinois Superintendents | Illinois Superintendents Large Schools | Ill. Supt. Medium Schools | Illinois Superintendents Small Schools |
| All Illinois Superintendents | -.48* | | | |
| Jury of Experts | -.51* | -.44* | -.39** | -.53* |
| Oklahoma Superintendents | -.43* | -.36** | -.32** | -.48* |
| Illinois Superintendents (Small Schools) | | | | -.42* |
| Illinois Superintendents (Medium Schools) | | | -.39** | |
| Illinois Superintendents (Large Schools) | | -.40** | | |

---

* Statistically significant at the 1% level of significance.
**Statistically significant at the 5% level of significance.

## TABLE V

### RELATIONSHIPS BETWEEN IMPORTANCE OF FUNCTIONS AND EXTENT TO WHICH ADMINISTRATORS DEVOTE "TOO LITTLE" TIME AND ENERGY ON FUNCTIONS

| Groups Determining Importance of Functions | Functions Which Receive "Too Little" Time and Energy by Group | | | | |
|---|---|---|---|---|---|
| | Oklahoma Supts. | All Illinois Supts. | Ill. Supts. Small Schools | Ill. Supts. Medium Schools | Ill. Supts. Large Schools |
| Jury of Experts | .46* | .55* | .51* | .39** | .54* |
| All Illinois Superintendents | | .52* | | | |

* Statistically significant at the 1% level of significance.
**Statistically significant at the 5% level of significance.

## TABLE VI

### RELATIONSHIPS BETWEEN IMPORTANCE OF FUNCTIONS AND EXTENT TO WHICH ADMINISTRATORS DEVOTE "TOO MUCH" TIME AND ENERGY ON FUNCTIONS

| Importance by Group | Functions Receiving "Too Much" | | | | |
|---|---|---|---|---|---|
| | Oklahoma Supts. | All Illinois Supts. | Ill. Supts. Small Schools | Ill. Supts. Medium Schools | Ill. Supts. Large Schools |
| Jury of Experts | -.53 | -.52 | -.47 | -.46 | -.31* |
| All Illinois Superintendents | | -.53 | | | |

*Significant at the 5% level; all other coefficients significant at the 1% level.

ships are found between what both Illinois superintendents and experts consider as the most important functions and those on which it is perceived that "too much" time is being spent. It appears that as school enrollment size increases, the frequency with which "too much" time and energy are expended on functions of lesser importance decreases. There appears to be no significant difference between the frequency with which Oklahoma and Illinois superintendents devote "too much" time and energy to functions of lesser importance.

Superintendents perceive the time and energy expended in regard to certain functions are "about right." Functions with respect to which the largest number of superintendents indicated their time and energy expenditure were "about right" are listed below in descending order of frequency with which they were mentioned:

1. Employing of clerks, custodians, other non-professional personnel.
2. Assigning, transferring, and dismissing personnel.
3. Keeping operation of school in proper legal channels.
4. Supervising attendance.
5. Preparing budget.
6. Maintaining working relations with PTA groups.
7. Interviewing patrons and representatives of local organizations.
8. Providing health examinations for pupils and employees.
9. Purchasing general supplies (non-educational).
10. Taking of school census.
11. Employing of teaching, supervisory, and administrative personnel.
12. Selecting library and other supplies.
13. Delegating authority.

More than three of every five superintendents perceived their practice as "about right" in time and energy expenditures for each of the functions listed. While the function of purchasing non-educational supplies was high among the functions on which administrators indicated they were devoting "too much" time and energy, it appears ninth in order among the functions in regard to which, in general, time and energy expenditures are indicated as being "about right." This is the only function of high rank order duplicated among the three listings, and all three listings of high rank functions contain all but six of the 38 functions. Thus it appears that there is distinct general agreement among participating Illinois superintendents concerning the evaluation of administrative practice relative to the various functions.

Relationship Between Relative Importance of Administrative Functions Based on Authoritative Judgment and Perceived Time and Energy Expenditures of Illinois Superintendents—Graphical pictures of the relationship between authoritative opinion on importance of functions and relative adequacy of time and energy expenditures for each function in two area classifications where the divergence was greatest are presented for Illinois superintendents. A similar picture was portrayed by Oklahoma administrators. This will reveal at a glance the extent to which practice regarding time and energy expenditures coincides with relative importance of items. Relative rank order percentile positions of relative importance were ascertained for each function, and they are indicated by the top bar opposite a function in each instance. A similar procedure relative to amount of time and energy expenditures was used as a basis for the second bar opposite each function. Graphical presentations of these relationships for other areas of function have by necessity been omitted due to space limitations.

Instruction—While, as indicated in an earlier portion of this report, experts considered the provision of leadership by administrators for the in-service training of the staff as being of higher relative importance than did superintendents, Figure 1 indicates relatively little time and energy are expended on this function compared to its position of importance. In general, this relationship is true for all functions related to the area of instruction. These data indicate that in terms of the importance of administrative functions related to the instructional program, school superintendents might well examine their present practices in the light of the total task with a view of increasing the time and energy expended on this vital area of functions.

Ministerial and Miscellaneous—As shown in Figures 2 and 3, compared to relative importance of specific functions in these two areas, a disproportionate amount of time and energy are expended on each function so classified. The most exorbitant expenditure of time and energy in relation to relative importance of functions is in relation to supervising attendance, designing record forms, completing reports to federal, state, and local authorities and administering lunch and other auxiliary programs. Keeping records, taking school census and administering the transportation program all receive a disproportionate amount of time and energy to a lesser degree.

Summary of the Findings—Data obtained in this study are summarized below to identify concisely the major findings:

1. City school superintendents in Oklahoma and Illinois superintendents of separate elementary school districts agree closely with each other

FIGURE 1

RELATIONSHIP BETWEEN IMPORTANCE OF FUNCTIONS RELATED TO INSTRUCTION AND TIME AND ENERGY EXPENDITURES

| | 0 10 20 30 40 50 60 70 80 90 100 |
|---|---|
| Providing leadership for in-service training of staff | xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx<br>** |
| Initiating curriculum study and revision | xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx<br>*********** |
| Supervising the improvement of instruction | xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx<br>(zero) |
| Holding and attending workshops, clinics, and institutes | xxxxxxxxxxxxxxxxxxxxxxxxxx<br>********** |
| Preparing courses of study | xxxxxxxxxxxxxxxxxxxxxxx<br>***** |

xxxx Percentile rank of relative importance.

**** Percentile rank of relative expenditure of time and energy.

FIGURE 2

RELATIONSHIPS BETWEEN MINISTERIAL FUNCTIONS AND TIME AND ENERGY EXPENDITURES

```
                                   0  10  20  30  40  50  60  70  80  90  100
                                   ------------------------------------------
Keeping proper records             xxxxxxxxxxxxxxxxxxxxxxx
                                   **********************************************

Reporting to federal, state, and   xxxxxxxxxxxxxxxxxxxxx
  local authorities                ****************************************************

Taking school census               xxxxxxxxxx
                                   ********************************
```

xxxx Percentile rank of relative importance.
**** Percentile rank of relative expenditure of time and energy.

FIGURE 3

RELATIONSHIPS BETWEEN MISCELLANEOUS FUNCTIONS AND TIME AND ENERGY EXPENDITURES

```
                                   0  10  20  30  40  50  60  70  80  90  100
                                   ------------------------------------------
Administering the transportation   xxxxxxxxxxxx
  program                          *****************************

Designing of record forms          xxxxxxxx
                                   *************************************

Supervising attendance             xxxxxx
                                   ***********************************

Administering lunch and other      xx
  auxiliary programs               **************************************************
```

xxxx Percentile rank of relative importance.
**** Percentile rank of relative expenditure of time and energy.

and with the opinions of authorities in the field of school administration about the relative importance of various administrative functions. However, Oklahoma superintendents agreed more closely with authoritative opinion on relative importance of various administrative functions of a superintendent than did Illinois separate elementary school district superintendents.

2. As school district enrollment increases, opinions of school superintendents in both Illinois and Oklahoma were found to agree more closely with those of authorities on relative importance of administrative functions.

3. Both groups of superintendents and the experts concur that administrative functions related to the board of education are of greatest relative importance or "extremely important" and that functions related to personnel administration are "significantly important." The experts consider public relations functions to be of higher relative importance than do either group of superintendents.

4. Superintendents concur in associating a higher degree of relative importance to functions related to pupils, finance, certain miscellaneous functions, and ministerial functions than do the experts.

5. The two areas of greatest discrepancy between experts and Illinois superintendents are public relations and pupil personnel administration. Experts associated greater relative importance than did superintendents with public relations functions, and the converse was true of pupil personnel administration.

6. The experts associated a higher relative degree of importance with the superintendent's leadership function relating to staff improvement than superintendents, and superintendents associated a higher relative degree of importance than experts on functions directly involving them in personally bringing about curriculum improvement through initiating curriculum study and revision, preparing courses of study, and supervising the improvement of instruction.

7. As perceived by both groups of superintendents, relatively too little time and energy are devoted to administrative functions which both they and the experts agree are of greatest relative importance and too much time and energy are expended on functions of lesser importance. These conditions are more true for Illinois superintendents of schools of small enrollments than for those of medium or large size.

8. Superintendents of medium size separate Illinois elementary school districts perceive they devote relatively more time and energy to functions deemed to be of greatest relative importance than superintendents in districts of either small or large size enrollments perceive they devote to such functions.

9. Half of the ten top ranked functions on which Illinois superintendents perceive they are devoting relatively "too little" time and energy pertain to administration of the instructional program. Another such function relates to the administrator's function of improving his own personal and professional qualifications. Another pertains to reporting to the public and two, formulating school policies and evaluating school progress, are functions relating to the Board of Education.

10. Illinois elementary school district superintendents, more than Oklahoma superintendents perceive they devoted "too little" time and energy to functions deemed by the jury and themselves as being of greatest relative importance, but there appeared to be no such difference between the two groups of superintendents concerning items on which relatively "too much" time and energy are expended in terms of relative importance of functions.

11. Functions concerning which Illinois superintendents perceive they are devoting relatively "too much" time and energy are primarily ministerial and miscellaneous in nature. These functions include those of keeping records, purchasing non-educational supplies and supervising non-teaching employees, designing record forms, completing reports to various governmental authorities, supervising school plant maintenance, dealing with pupil problems and administering lunch, transportation, and other auxiliary programs.

12. Comparative analysis of percentile rank positions of function importance and relative time and energy expenditures of Illinois superintendents reveals that functions related to instruction are the ones where there is the greatest consistent expenditure of relatively "too little" time and energy in terms of relative importance accorded functions by the experts. This is generally true of functions related to the Board of Education.

13. Comparative analysis of percentile rank position of function importance and relative time and energy expenditures of Illinois superintendents reveals the ministerial and miscellaneous area of functions to be those where there is the greatest consistent expenditure of

relatively "too much" time and energy in terms of relative importance accorded functions by the experts.

Conclusions and Implications—Based upon the findings of this study, the investigators have reached certain conclusions and drawn pertinent implications as follows:

1. Perceptions of importance held by Illinois elementary school superintendents regarding relative importance of administrative functions substantially agree with those held by experts in the field of school administration.

2. Superintendents of schools know which of their functions are of most importance, but for some reason in practice they are either thwarted from devoting greatest attention to the most important aspects of the task of school superintendent, tending without realizing it, to feel more secure in concerning themselves with functions requiring less educational statesmanship, or by choice are reluctant to concern themselves with tasks of greatest importance which involve higher degrees of risk in successfully accomplishing them. Further study is needed to reveal the causal factors.

3. There is a statistically significant inverse relationship between functions accorded greatest importance by Illinois elementary school superintendents and their perceived time and energy expenditures relative to those functions. This is true also concerning what authorities consider as important.

4. The validity of time and energy expenditures of Illinois separate elementary school district superintendents decreases as school enrollment size diminishes, based on authoritative opinion.

5. Increased divergence of agreement on the relative importance of a school superintendent's functions between experts and superintendents and validity of time and energy expenditures which are associated with size of school may be attributed to the more generalized nature of administrative function as school size decreases, to the lesser amounts and quality of training and experience of persons employed in smaller schools, or possibly to the fact that relative importance of functions may actually vary according to enrollment size of the school.

6. The validity of perceived time and energy expenditures among Illinois administrators included in the study is less than for Oklahoma school superintendents based upon authoritative opinion, particularly with reference to time and energy expenditures related to functions considered of greatest relative importance, namely, the area of instruction.

7. Illinois superintendents of separate elementary school districts could profitably re-examine the manner in which they are discharging the task of school administration particularly with reference to their functions relating to the Board of Education and administration of the instructional program.

8. Not only do the results of this study point to the desirability for practicing Illinois superintendents to re-examine their practices in regard to administrative functions, but they also have implications for the content and emphases found in college or university administrator preparation programs.

9. Based on authoritative judgment, elementary school superintendents could profitably reconsider the importance of public relations functions and their role as administrator in the improvement of instruction and the curriculum.

10. The writers conclude that similar findings would have been obtained if every separate elementary school district superintendent in Illinois had participated in the study, and they suspect the picture would be little different if superintendents of unit or separate secondary school districts had been included. Improvement of educational administrative practice is the responsibility of both the practicing administrator and the researcher who may probe more deeply into the obstacles and causal factors surrounding the condition where greatest attention is devoted to functions of lesser importance with a view to re-orienting emphasis of effort.

11. The findings of this study precipitate questions concerning basic characteristics of the structure within which the system of public education is provided. Is there a "best" size of school organization within which the chief executive is most able to function as a professional leader in accomplishing the task? Is it possible that the basic organizational structure, within which the school exists, can facilitate or hinder the most effective performance of the chief executive in terms of the real purpose for which public schools have been established? Theoretically, assuming that three superintendents with identical training, experience, competencies, abilities, and motivations secured positions in large, medi-

um, or small size school systems, respectively, in communities with identical characteristics, is it possible that with equal effort one can be more effective than the other two in discharging this task of educational leadership because of more favorable conditions associated with differences in basic organizational structure?

## FOOTNOTES

1. J. K. Norton, "Meeting the Challenges to Educational Administration," Teachers College Record, LII (May 1951), pp. 467-8.
2. American Association of School Administrators, Yearbook (Washington, D. C., 1933), pp. 135-42.
3. American Association of School Administrators, The American School Superintendency, 13th Yearbook (Washington, D. C.: The Association, 1952), p. 293.
4. R. R. Russell, A Study of the Relationship Between the Importance of Administrative Functions and Time Expenditure of School Superintendents, Based on Authoritative Judgment, unpublished Ph. D. dissertation, Oklahoma A and M College, Stillwater, 1954.
5. Ibid., p. 77.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 28, Number 4, June 1960)

# A COMPARATIVE STUDY OF STUDENT-TEACHING PRACTICES IN THIRTY-EIGHT MIDWEST INSTITUTIONS

GAIL M. INLOW
Northwestern University

## Introduction—Background and Sample

AS ONE MEANS of evaluating its program of student teaching, the School of Education of Northwestern University recently became interested in analyzing student-teaching programs of other selected colleges and universities in the midwest area. Accordingly, in December of 1957, a questionnaire was mailed to the other Big Ten Universities and to 43 additional institutions of higher learning in the states of Illinois and Indiana. The 43 deans, chairmen of departments of education, and directors of student teaching whose names appeared in the *Thirty-Sixth Yearbook*, 1957, of the Association for Student Teaching, and the deans or directors of student teaching of the Big Ten Universities comprised the mailing sample.

Of the 52 institutions which received questionnaires, replies were returned by 39, or 75 percent. One reply was too incomplete to use, leaving a working sample of 38. Returns were as follows: 8 of the 9 Big Ten Universities returned questionnaires, 10 of the 11 other universities, all 5 of the teachers colleges (3 state and 2 private), and 15 of the 27 liberal-arts colleges. The respondent sample consists of 18 universities, 5 teachers colleges, and 15 liberal-arts colleges. Of the 27 institutions in Illinois which were contacted, 21 returned completed questionnaires; of the 16 in Indiana, 10 made returns. From this point on in the study, no distinction will be made among the institutions on the basis of location by state.

Enrollments for the year 1957-58 in the 38 institutions were listed as follows: 15 thousand or more, 6 institutions; 11-13 thousand, 2; 7-9 thousand, 1; 5-7 thousand, 4; 3-5 thousand, 4; 2-3 thousand, 3; 1-2 thousand, 3; 500-1 thousand, 12; and under 500, 3.

Mean average student teaching enrollments on hand and projected for the remainder of the year 1957-58 were listed as follows: 250 elementary and 430 secondary for the Big Ten schools, 86 elementary and 176 secondary for the other universities, 174 elementary and 247 secondary for the liberal-arts colleges, and 113 elementary and 195 secondary for all colleges and universities. The total for all institutions was 4078 elementary student teachers and 6421 secondary.

All but one of the 38 institutions, a liberal-arts college, had a program of elementary education; all but two, both teachers colleges, had a program of secondary education.

Ten of the 18 universities, 4 of the 5 teachers colleges, and 1 of the 15 liberal-arts colleges had laboratory schools at the elementary level; 10 of the 18 universities, 2 of the 5 teachers colleges, and 4 of the 15 colleges had laboratory schools at the secondary level. Totals were 15 for elementary and 14 for secondary.

## The Purpose

The purpose of the study was to discover and evaluate existing practices in the program of student teaching of other institutions of higher learning to the possible and of making changes in the program of the School of Education of Northwestern University in areas where changes seemed warranted. In brief, the study was intended as a stimulus for a more effective program evaluation. Information was requested in an effort to determine comparative practices in the following facets of the student-teaching program.

1. How many credit hours of student teaching are required or permitted by the colleges and universities which responded to the questionnaire?
2. What basic patterns of student teaching are followed: hours per day, days per week, and number of weeks?
3. How many secondary-school students have experiences in both their major and minor fields of subject matter?
4. What kinds of compensation are made to public and private-school personnel who cooperate in the program?
5. How are cooperating teachers selected?
6. How adequate are cooperating teachers believed to be?
7. How are students screened prior to accept-

ance for student teaching?

8. What practices are followed prior to student teaching in getting students acquainted with children and youth?

Data pertaining to the role of the college supervisor and an overall evaluation of the 38 programs of student teaching will be published at a later date. The length of the questionnaire and complexity of some of the returns led the investigator to report the findings in several rather than one manuscript.

## The Findings of the Study

Credit hours of student teaching. One of the current problems of teacher education is to determine how many hours of student teaching should be required. To discover existing practices among the various colleges and universities, the investigator included the following item in the questionnaire:

> "Within the framework of an undergraduate-degree program, please list in the following table the number of credit hours of student teaching required for graduation. If additional hours of student-teaching enrollment are permitted, please write in the number of hours for each year of school as indicated. If student-teaching is a senior-year experience for a majority of students, but a small minority are enrolled in it during the junior year, list it in the senior year only."

The data which were received in response to the item are presented in Table I.

The most distinguishing feature of the data is their variability. The required hours of student teaching have a range from 5 to 21 quarter hours in elementary and 5 to 15 hours in secondary. The Big Ten Universities are at the top in elementary and secondary for both mean and median hours, followed in rank order by the teachers colleges, other universities, and liberal-arts colleges. The School of Education of Northwestern University requires 16 quarter hours of student teaching at both the elementary and secondary levels.

In addition to the required hours listed in Table I, the mean number permitted on an optional basis is 2.11 quarter hours at the elementary and 2.26 at the secondary level. The Big Ten Universities permit on an optional basis 1.74 quarter hours at the elementary and 1.69 at the secondary level; other universities, 1.90 and 3.10 respectively; teachers colleges, 3.40 and 2.66; and liberal-arts colleges, 2.03 and 1.90.

When mean required and additional hours are combined, it may be observed that at the elementary level, the mean averages are as follows: Big Ten, 14.81 quarter hours; other universities, 13.45; teachers colleges, 15.50; and liberal-arts colleges, 10.85. At the secondary level, totals are 12.00, 12.50, 14.49, and 9.97 respectively.

Hourly, daily and weekly pattern of student teaching. Of interest to the investigator was the time pattern which the various institutions followed in their student-teaching programs. Data were returned as presented in Table II.

It is to be noted from the last line of Table II that the dominant pattern of student teaching at both the elementary and secondary levels is a half-day experience: 19 of the 38 institutions at the elementary and 15 of the 36 at the secondary level follow this pattern. Next in quantitative importance is the full-day program with totals of 9 and 8 respectively. Six at the elementary level follow both patterns; 6 and 2 respectively are on a 2 hour per day basis; 3 secondary on a 1 hour basis; and 1 secondary on a 3/4 hour basis. All but two institutions have student teaching five days weekly.

The range in terms of weeks of student teaching is varied depending on whether the institution is on a full-day, half-day, or some other organizational basis. The range at the elementary level is 9-27 weeks; at the secondary, 8-32. The range for the full-day program is 6 to 14 weeks with only one university on a 6 week plan; all others were on a 9 to 14 week plan.

Of interest is the fact that only minor organizational differences exist between the elementary and the secondary programs in terms of hours per day, days per week, and number of weeks of student teaching. The program of the school of Education of Northwestern University reflects the same minor differences. The basic pattern is a full-day program for 11 weeks, 5 days per week. However, more full-day students proportionately are accepted in the secondary than in the elementary program. The impact of an inadequate full-day student teacher at the elementary level in a self-contained classroom is so great that many students are encouraged to enroll in the half-day program for two quarters.

Student teaching in the major and minor fields. One of the controversial areas in the field of secondary education is whether students should be graduated with teaching competencies in one field or in several; and if in several, whether these competencies should be demonstrated in a student-teaching situation. The questionnaire item which related to this problem elicited responses as indicated in Table III.

An analysis of Table III indicates that the tendency of most student teachers, 68.21%, who are enrolled in the 36 institutions which have secondary programs is to teach only in their major fields. The liberal-arts colleges lean most heavily in the direction of specialization whereas the Big Ten schools are the least prone to narrow student teaching horizons to only a major area of concentration. By the same logic, 45.86% of the secondary students in Big Ten schools have a practical experience in both the major and minor fields whereas only 17.8 % of the student teachers in liberal-arts colleges have a similar experience.

TABLE I

QUARTER HOURS OF STUDENT TEACHING

| School | Elementary | | | Secondary | | |
|---|---|---|---|---|---|---|
| | Mean Reqs. | Median Reqs. | Range | Mean Reqs. | Median Reqs. | Range |
| Big Ten Universities | 13.07 | 12.33 | 7.5-19 | 10.31 | 10.5 | 7.5-15 |
| Other Universities | 11.55 | 11.75 | 5.21 | 9.4 | 9.75 | 5-12 |
| All Universities | 12.23 | 11.95 | 5.21 | 9.81 | 10.10 | 5-15 |
| Teachers Colleges | 12.10 | 12.00 | 8-15 | 9.83 | 12.00 | 8-13.5 |
| Liberal-Arts Colleges | 8.82 | 7.35 | 7-15 | 8.07 | 7.01 | 6-15 |
| All Colleges | 9.60 | 7.90 | 7-15 | 8.47 | 9.55 | 6-15 |
| All Institutions | 10.83 | 10.15 | 5-21 | 9.09 | 9.80 | 5-15 |

TABLE II

TIME PATTERN FOR STUDENT TEACHING

| Schools | Hours per Day | | Days per Week | | No. of Weeks | |
|---|---|---|---|---|---|---|
| | Elem. | Second. | Elem. | Second. | Elem. | Second. |
| Big Ten Universities | 3 f. d.<br>4 h. d.<br>1 No Ans. | 3 f. d.<br>2 h. d.<br>1-2 hr.<br>1-1 hr.<br>1 No Ans. | All 5 d. | All 5 d. | 10-20 | 8-32 |
| Other Universities | 5 f. d.<br>5 h. d.<br>2 b. | 3 f. d.<br>5 h. d.<br>1-2 hr.<br>1-3/4 hr. | All 5 d. | All 5 d. | 9-24 | 9-18 |
| All Universities | 6 f. d.<br>9 h. d.<br>2 b.<br>1 No Ans. | 6 f. d.<br>7 h. d.<br>2-2 hr.<br>1-1 hr.<br>1-3/4 hr.<br>1 No Ans. | All 5 d. | All 5 d. | 9-25 | 8-32 |
| Teachers Colleges | 1 f. d.<br>1 h. d.<br>3 b. | 1 f. d.<br>2 h. d.<br>2 No Pgm. | 1-4 1/2 d.<br>1-3 d.<br>3-5 d. | 1-2 d.<br>2-5 d. | 12-27 | 8-19 |
| Liberal-Arts Colleges | 2 f. d.<br>9 h. d.<br>1 b.<br>2-2 hr.<br>1 No Ans. | 1 f. d.<br>6 h. d.<br>4-2 hr.<br>2-1 hr.<br>2 No Ans. | All 5 d. | All 5 d. | 9-18 | 9-18 |
| All Colleges | 3 f. d.<br>10 h. d.<br>4 b.<br>2-2 hr.<br>1 No. Ans. | 2 f. d.<br>8 h. d.<br>4-2 hr.<br>2-1 hr.<br>2 No Ans. | 18-5 d.<br>1-4 1/2 d.<br>1-3 d. | 17-5 d.<br>1-2 d.<br>2 No Pgm. | 9-27 | 8-19 |
| All Institutions | 9 f. d.<br>19 h. d.<br>6 b.<br>2-2 hr.<br>2 No Ans. | 8 f. d.<br>15 h. d.<br>6-2 hr.<br>3-1 hr.<br>1-3/4 hr.<br>5 No Ans. | 36-5 d.<br>1-4 1/2 d.<br>1-3 d. | 35-5 d.<br>1-2 d.<br>2 No Pgm. | 9-27 | 8-32 |

Legend: f. d. = full day; h. d. = half day; b. = both full day and half day; No Pgm.= No Program.

## TABLE III

## STUDENT TEACHING IN THE MAJOR AND MINOR FIELDS

| Schools | Major Only | Minor Only | Both |
|---|---|---|---|
| Big Ten Universities | 52.00% | 2.14% | 45.86% |
| Other Universities | 63.89 | 1.22 | 34.89 |
| All Universities | 58.69 | 1.63 | 39.68 |
| Teachers Colleges | 68.33 | 1.67 | 30.00 |
| Liberal-Arts Colleges | 78.37 | 3.83 | 17.80 |
| All Colleges | 76.67 | 3.50 | 19.83 |
| All Institutions | 68.21 | 2.61 | 29.18 |

At Northwestern University, student teachers who are under the auspices of the school of Education have experiences in both the major and minor fields, with the exception of students in the areas of art, home economics and business education. In general, members of the staff of the School of Education favor breadth of teaching preparation; of the staff of the liberal-arts college, more specialized preparation.

Compensation to the cooperating schools. How many colleges and universities pay teachers or school systems for the services which they render the student teaching program? How many, and in what way, do institutions compensate other than by monetary payments to teachers or school systems? Answers were sought to these and other related questions about compensation for student teaching services which are performed by those public and private schools which cooperate in the teacher-education programs of the 38 respondent institutions. Pertinent findings are contained in Table IV.

Of the 38 institutions which returned usable questionnaires, 28 indicated that they made direct payments to classroom teachers for services rendered. As indicated in line 1 of Table IV, the mean average payments per quarter hour of student teaching time ranged from $1.33 to $11.00 with a mean average of $6.06 for all institutions. The median was $5.08. The Universities paid an average of $6.81 and the colleges, $5.31.

The mean institution of the sample with a student teaching enrollment of 113 elementary and 195 secondary students scheduled for a mean of 10.83 and 9.09 required hours of student teaching respectively would pay to cooperating teachers a total of $18,157 per year.

Two schools, both teachers colleges, make a payment to the school system, not to the teachers themselves. These two along with 28 that make a direct payment to teachers leave only 8 of the 38 institutions which do not pay either teachers or the school board.

As may be noted in Table IV, 6 institutions gave tuition benefits of some kind to the cooperating teachers or schools; 20 extended library privileges; 10 permitted cooperating teachers to attend special events at the college or university, and 10 listed the names of teachers in the college catalog. One institution which makes no direct payments to teachers or a school board stated that it had $90,000 to $120,000 per year for tuition benefits to teachers in the cooperating school systems for in-service purposes, workshops and professional meetings.

Other miscellaneous methods of compensation were listed as follows, with parenthetical figures indicating the number of institutions which used the specific medium: pay principal (1), give dinners (4), give teas (2), make union facilities available (1), provide equipment (2), subscribe to professional journals for teachers and schools (2), finance workshops (2), invite teachers to staff meetings (1), and charge a tuition fee of $10.00 per student which goes to the teacher (1).

Northwestern University makes no direct payment to teachers but extends limited tuition benefits, gives faculty rates to teachers for special events, extends library privileges, and includes cooperating personnel on its mailing list.

How are cooperating teachers selected? A problem that regularly seems to plague teacher-education institutions is how to discover before the fact whether a teacher is likely to be an effective cooperating teacher. To get acquainted with comparative practices, the following item was included in the questionnaire:

What methods do you use in selecting cooperating teachers? After each method listed, please indicate how effective the method is by placing a check under one of the four listed symbols: V. E. for very effective, M. E. for moderately effective, M.I. for moderately ineffective, and V.I. for very ineffective.

Responses to the item are presented in Table V. Symbols are employed as indicated above.

From the data in Table V, no conclusions may be reached about the best way or ways of selecting supervising teachers. The majority of institutions, 31, seem to permit school officials to help select cooperating teachers; 24 visit classrooms; and a small minority use other methods. Northwestern University is likewise perplexed over the problem of how to select capable cooperating teachers. Apparently only after an institution has worked with a teacher for an extended period of time can it be reasonably certain about the accuracy of the original selection.

How competent are supervisi [illegible] Did the 38 institutions believe that they were [illegible] with reasonably competent classroom teachers in their student teaching programs? The data in Table VI, at least in part, provide the answers.

The implications of Table VI are encouraging. In the opinion of the respondents, a vast majority of the cooperating teachers are believed to meet high standards or to do an adequate job. Interesting is the fact that the declared adequacy of the classroom teacher seems to vary but little whether he works with a student teacher from a university, a teachers college, or a liberal-arts college. Also of interest is the fact that little difference exists between the declared competency of elementary and secondary-school teachers. The over-all picture is as follows: in the elementary field, 92.7 percent were believed to meet high standards regularly or to do an adequate job; in the secondary field, 91.6 percent were identified as meeting these same standards. The Big Ten Universities were the most pessimistic, labeling 12.2 percent of the secondary school teachers as less than ade-

## TABLE IV

## COMPENSATION TO COOPERATING SCHOOLS AND PERSONNEL

| | Big Ten | Other Uni. | All Uni. | Teachers Colleges | Lib.-Arts Colleges | All Colleges | All Inst. |
|---|---|---|---|---|---|---|---|
| Direct payment per quarter hour of student teaching. | $\$5.61^{6}$<br>(4.16-<br>7.50) | $\$7.83^{7}$<br>(5.00-<br>11.00) | $6.81 | $\$8.50^{2}$<br>(7.00-<br>10.00) | $\$4.82^{13}$<br>(1.33-<br>10.00) | $5.31 | $\$6.06^{28}$ |
| No. of schools which extend library privileges. | 4 | 3 | 7 | 5 | 8 | 13 | 20 |
| No. of schools which give tuition benefits. | 3 | 0 | 3 | 0 | 3 | 3 | 6 |
| No. of schools which permit teachers to attend special activities at faculty rates. | 1 | 2 | 3 | 1 | 6 | 7 | 10 |
| No. of schools which list co-operating teachers' names in college catalog. | 1 | 3 | 4 | 3 | 3 | 6 | 10 |

Note: Line one is to be interpreted as follows: $5.61, for instance, represents the mean payment made by Big Ten schools per quarter hour of student teaching; the exponents 6, 7, 2, 13 and 28 tell the number of Big Ten schools which pay; and the numbers in parentheses give the range.

TABLE V

METHODS OF SELECTING SUPERVISING TEACHERS

| | Big Ten | Other Uni. | All Uni. | Teachers Colleges | Lib.-Arts Colleges | All Colleges | All Inst. |
|---|---|---|---|---|---|---|---|
| Selected directly by cooperating school officials. | 3 VE<br>4 ME<br>1 MI | 2 VE<br>5 ME | 5 VE<br>9 ME<br>1 MI | 1 VE<br>3 ME | 8 VE<br>4 ME | 9 VE<br>7 ME | 14 VE<br>16 ME<br>1 MI |
| University personnel visit classrooms. | 3 VE<br>4 ME | 5 VE<br>1 ME | 8 VE<br>5 ME | 5 VE | 5 VE<br>1 ME | 10 VE<br>1 ME | 18 VE<br>6 ME |
| Recommended by colleagues, teachers. | | | | 1 ME | 1 VE<br>1 ME | 1 VE<br>2 ME | 1 VE<br>2 MI |
| Follow policies and standards of state department. | 2 ME | | 2 ME | 1 ME<br>1 MI | 1 ME | 2 ME<br>1 MI | 4 ME<br>1 MI |
| Use teachers who volunteer | 1 VE | 1 ME | 1 VE<br>1 ME | 1VE | 1 VE | 2 VE | 3 VE<br>1 ME |
| Use teachers who are recommended by students. | | | | 1 VE<br>2 ME | 1 MI | 1 VE<br>2 ME<br>1 MI | 1 VE<br>2 ME<br>1 MI |

Legend: VE = very effective; ME = moderately effective; MI = moderately ineffective; VI = very ineffective.

## TABLE VI

## HOW COMPETENT SUPERVISING TEACHERS ARE BELIEVED TO BE

| | | Big Ten | Other Uni. | All Uni. | Teachers Coll. | Lib. Arts Coll. | All Coll. | All Inst. |
|---|---|---|---|---|---|---|---|---|
| Almost always meet high standards | Elem. | 58.6*<br>(30-80) | 66.0<br>(5-95) | 62.9 | 59.2<br>(25-99) | 64.6<br>(0 95) | 63.0 | 63.0 |
| | Second. | 60.0<br>(10-85) | 66.5<br>(5-95) | 63.8 | 56.3<br>(20-99) | 58.5<br>(0-100) | 58.0 | 61.0 |
| Do an adequate job | Elem. | 29.1<br>(15-50) | 29.5<br>(5-90) | 29.4 | 33.2<br>(0-60) | 29.9<br>(5-100) | 30.0 | 29.7 |
| | Second. | 26.3<br>(10-60) | 28.5<br>(5-95) | 27.4 | 33.7<br>(0-40) | 34.4<br>(10-100) | 34.2 | 30.6 |
| Are borderline cases | Elem. | 11.0<br>(0-25) | 4.4<br>(0-10) | 7.1 | 4.6<br>(0-10) | 4.6<br>(0 40) | 4.6 | 6.3 |
| | Second. | 11.0<br>(0-25) | 4.5<br>(0-15) | 7.3 | 7.3<br>(0-15) | 6.3<br>(0-35) | 6.5 | 7.0 |
| Should not be used | Elem. | 1.30 | 0.1 | 0.6 | 3.0 | 0.9 | 1.5 | 1.0 |
| | Second. | 2.7 | 0.5 | 1.5 | 2.7 | 0.8 | 1.3 | 1.4 |

* The figure of 58.6 is the mean percentage; the parenthetical figures represent the range of percents.

quate--either borderline cases or persons who should not again be used.

In the opinion of the investigator, the percentages subscribed to by the Big Ten Universities are fairly realistic. If he differed at all, he would lean somewhat more to the pessimistic than to the optimistic side. He believes that securing the services of enough top-notch teachers is one of the big problems in teacher education today.

The screening of student-teaching applicants. In the area of screening for student teaching, the institutions were presented with a list of nine methods or techniques and asked to respond with a symbol R if they routinely used the method; W if they used it when warranted; and N if they used it rarely or never. Only the responses of R or W have been included in Table VII.

As an analysis of Table VII indicates, many and varied are the approaches of the participating institutions to the problem of screening applicants. They agreed unanimously on only one method--the academic record. Next in order of support as part of their routine procedures were the student-teaching interview (29), the physical exam (19), speech test (19), standardized instruments (15), autobiography (15), written faculty references(14), written character references (12), and psychiatric exam (4).

Once more the fact is significant that the number of methods and devices employed are not related significantly to the type of institution. The Big Ten led the field in the areas of speech testing and the use of standardized instruments of various kinds. The colleges, on the other hand, were more inclined to secure professional and character references as part of their screening procedures.

As indicated in Table VII, 15 institutions employed standardized instruments for screening purposes: academic aptitude tests, English Usage tests, reading tests, personality inventories, and attitude inventories. The specific instruments listed were the ACE, SCAT, Miller Analogy, Ohio Psychological, Otis, Kuder personality, M. M. P. I., and the Minnesota Teacher Attitude Scale.

Northwestern University regularly requires a physical exam, student-teaching interview, professional references, and an acceptable academic record. When warranted, it requires a psychiatric exam, speech test, and an occasional standardized test in English usage.

How students, prior to student teaching, are helped to understand children and youth. An item was included in the questionnaire in an effort to discover practices which other institutions employed as a pre-student-teaching experience to help students achieve a better understanding of children and youth. Specifically it asked:

"If your college or university has no organized program of student-teaching prior to the senior year, what procedures, if any, are used to help students discover the interests, abilities, attention spans, range of normalcy and other characteristics of children and youth; and the leadership traits and interests of the students themselves?"

Three of the 38 institutions seemed to have a well organized program of observation and student-teaching participation which begins prior to the senior year. One of the three, a university, requires for both elementary and secondary students three semester hours of observation in each of the freshman and sophomore years, permits five semester hours of student teaching in the junior year, and requires five semester hours of student teaching during the senior year. A teachers' college requires 4 quarter hours of student teaching in the elementary program during the junior year and requires 8 more in the senior year with 4 additional permitted. A second teachers' college requires for elementary students 3 semester hours of student teaching during the sophomore year, 6 during the junior year, and permits 6 additional during the senior year.

As a means of getting students acquainted with children and youth, probably second only to student teaching itself, is the making of visits to classrooms. Respondent sample members indicated that visits by students were made in the manner indicated below. The number of colleges and universities which utilize each method is listed in the parentheses which follow each item.

1. Visit in connection with courses in educational psychology and child development (10).
2. Visit in connection with methods courses(7).
3. Visit in connection with courses in educational sociology and human development (2).
4. Visit in connection with a course in directed observation (7).
5. Visit with an entire class (1).
6. Make frequent observations (10).

The 36 institutions made 37 references to classroom visits. One of the 10 institutions which admitted to making frequent observations stated that 20 to 25 were made by each student in the junior year. Another stated that 10 observations were required in each course in educational psychology and general psychology.

Next to visitations to schools, in order of frequency of mention, was youth work in service agencies. Because some doubt existed in the mind of the investigator about whether some of these experiences were required or optional, exact replies are provided as follows:

1. Youth activity required as part of junior sequence (for both elementary and secondary).(Big Ten University)
2. Leadership experience in community service agencies. (Big Ten University)
3. Non-teaching experience with children. (Other than Big Ten University)
4. Instituted this year--preteaching club activ-

## TABLE VII

## METHODS AND TECHNIQUES OF SCREENING APPLICANTS

| Screening Devices | Big Ten | Other Uni. | All Uni. | Teachers Colleges | Lib.-Arts Colleges | All Colleges | All Inst. |
|---|---|---|---|---|---|---|---|
| Physical Exam. | 2-R<br>3-W | 8-R<br>0-W | 10-R<br>3-W | 0-R<br>0-W | 9-R<br>2-W | 9-R<br>2-W | 19-R<br>5-W |
| Psychiatric Exam. | 1-R<br>2-W | 1-R<br>2-W | 2-R<br>4-W | 0-R<br>0-W | 2-R<br>0-W | 2-R<br>0-W | 4-R<br>4-W |
| Student-Teaching Interview | 6-R<br>2-W | 8-R<br>1-W | 14-R<br>3-W | 2-R<br>1-W | 13-R<br>1-W | 15-R<br>2-W | 29-R<br>5-W |
| Speech Test | 6-R<br>1-W | 5-R<br>2-W | 11-R<br>3-W | 4-R<br>0-W | 4-R<br>6-W | 8-R<br>6-W | 19-R<br>9-W |
| Academic Record | 8-R | 10-R | 18-R | 5-R | 15-R | 20-R | 38-R |
| Written Faculty References | 2-R<br>2-W | 4-R<br>4-W | 6-R<br>6-W | 1-R<br>2-W | 7-R<br>2-W | 8-R<br>4-W | 14-R<br>10-W |
| Written Character References | 0-R<br>4-W | 4-R<br>2-W | 4-R<br>6-W | 0-R<br>0-W | 8-R<br>4-W | 8-R<br>4-W | 12-R<br>10-W |
| Autobiography | 3-R<br>1-W | 4-R<br>1-W | 7-R<br>2-W | 2-R<br>0-W | 6-R<br>0-W | 8-R<br>0-W | 15-R<br>2-W |
| Standardized Tests and Inventories | 5-R<br>1-W | 2-R<br>0-W | 7-R<br>1-W | 0-R<br>0-W | 7-R<br>0-W | 7-R<br>0-W | 14-R<br>1-W |

Legend: R = method used routinely; W = method used only when warranted.

ities at freshman and sophomore and junior levels. (Other than Big Ten University)

5. Extended experience--two semesters of work in an approved social agency (junior year). (Other than Big Ten University)

6. Students are encouraged to serve as recreational leaders, etc., for groups of children. (Teachers College)

7. Guidance is provided toward getting as broad experience as possible with children and youth in church, and all types of youth organizations--also summer camps. (Liberal-Arts College)

8. Recommendation for participation with youth groups such as scouts, Sunday school camps. (Liberal-Arts College)

9. Directed leadership experience with Youth Day camps, summer camps. (Liberal Arts College)

10. Sunday school work, camp counseling, playground work. (Liberal-Arts College)

11. Playground work in summer camp counseling... teaching Sunday school classes. (Liberal Arts College)

12. Experiences with children in and out of school groups--Boy and Girl Scouts, Sunday school classes, etc.... summer camp experiences. (Liberal-Arts College)

Of the 36 institutions which replied to the item on pre-student-teaching activities, 12 or one-third described as optional or required selected activities with youth in various types of service agencies. Five of these institutions were universities, one a teachers college, and 6, liberal-arts colleges. On the negative end of the continuum was a cryptic reply "courses"; another, "none other than class work."

In addition to routine classroom observation and work in service agencies, other ways of getting students acquainted with children and youth were listed as follows:

1. September experience (4)
2. Future Teachers of America meetings (2)
3. Teacher-aide activity (1)
4. Conferences (1)
5. Resource persons from local community(1)
6. Classes taught by laboratory school and theory instructors (1)
7. Orientation tests and inventories (1)

In the opinion of the investigator, institutions, with only 3 or 4 exceptions, revealed an interest in pre-student-teaching experiences either by describing activities performed or by expressing regrets that more could not be done. What appears to be the most comprehensive reply came from a Big Ten University. It is as follows:

This phase of our teacher preparation needs the greatest strengthening. Most of us in student teaching feel that we are weak here. We do the usual sort of thing: observation, some visitation on a rather irregular basis, and the like. However, all students here must take two six-credit courses (core courses) prior to student teaching.

The School and the Child (psychologically oriented) and School and Society (orientation to school and community and to sociology of education). Individual professors probably do more than some others for students in this respect. Those that believe whole-heartedly in pre-laboratory experiences for students do a great deal; others do very little.

The School of Education of Northwestern University has two approaches to pre-student-teaching activities with children and youth: occasional observations in schools in connection with education courses (and these observations vary in number and quality dependent on the individual instructor), and a requirement of 100 hours of work with children and youth. These 100 hours may be spent in camp counseling, community centers, scouting organizations, YMCA's, and similar type organizations.

## Summary

In an effort to secure information which might be used in evaluating the program of the school of Education of Northwestern University, a questionnaire was sent to 52 teacher-training institutions to get an overview of comparative practices. Nine of these were Big Ten Universities; the other 43 were colleges and universities in the states of Illinois and Indiana, exclusive of the three Big Ten schools in those states. Replies were received from 39 or 75 percent; 38 were usable.

The 38 institutions had a mean average student-teaching enrollment of 113 elementary and 195 secondary students per year. The elementary students taught an average of 10.83 quarter hours with 1.74 additional hours optional; the secondary 9.09 with 3.40 hours optional. The predominant organizational pattern for both elementary and secondary was the half-day program of student teaching, followed by the full-day program, and a few institutions with 1 or 2-hour programs per day. The full-day students taught for a period of from 6 to 14 weeks; all but those of one institution from 9 to 14. The range for the remainder was 8-32 weeks.

A majority of the students, 68.21 percent at the secondary level, taught only in their major field of concentration; 29.18 percent taught in both the major and minor fields. The Big Ten schools led the way with 45.86 percent of their students teaching in both fields.

Thirty institutions made payments directly to teachers (28), or to the cooperating school systems (2). The mean average payment to teachers was $6.06 per quarter hour of student teaching. Other means of compensation were in tuition benefits, faculty rates to college events, library privileges, and other incidental media.

Cooperating teachers were shown to be selected in a variety of ways, with no one method being markedly superior to any other. Probably the best method was for college personnel to visit the teachers and classrooms, observe and get acquainted, and then place a student during a subsequent term. By and large, cooperating teachers were described as almost always meeting high standards or as doing an adequate job. Fewer than 10 percent were identified as borderline cases or in need of replacement.

Student teaching applicants were screened in many ways, with half or more of the respondents agreeing on the use of academic records, a physical examination, a speech test, and an interview as screening devices.

The two most commonly used methods of getting students acquainted with children and youth prior to the time of student teaching were listed as classroom visitation to public and laboratory schools, and work with youth in service agencies. Four institutions employed the September-Experience technique.

## Conclusions

In attempting to relate the findings of the study to possible program revision, a number of tentative conclusions may be drawn:

1. Programs were earmarked as much by their variety as by their similarity. Ranges of data were so tremendous at times that medians were more appropriate measures of central tendency than means.

2. Yet within the framework of variety, a surprising similarity existed among the institutions of different kinds. The investigator got the impression that differences among certain liberal-arts colleges, for instance, were much greater than the differences among selected liberal-arts colleges, teachers colleges, and universities that had a comparable philosophy of teacher education.

3. With few exceptions, the thirty-eight institutions used the state requirement of 5 semester hours of education (6 for elementary in Minnesota) as a guide line only. The more important question seemed to be how little or how much it should be exceeded in the light of sound educational practice. Of importance is the fact that few significant differences were revealed between the time schedules which were employed in secondary and elementary programs.

4. The data are conclusive in indicating that payment of teachers is a common practice among the 38 respondent institutions. Of equal significance may be the fact that 8 of the respondent institutions do not pay and yet have acceptable programs of student teaching.

5. The institutions are basically satisfied with the cooperating teachers with whom they work. The job of the future may be to secure the services of other well qualified teachers so that the current glow of satisfaction will not disappear.

6. Little seems to be done in screening in the area of personality. The fact that 29 of the 38 institutions used the student-teacher interview as a screening device may be a move in the right direction. But more probably needs to be done in the future with clinical instruments and psychiatric methods.

7. The majority of the colleges and universities were actively engaged in providing laboratory experiences with children and youth as part of the pre-student-teaching phase of the program.

In the last analysis, program revision will depend more emphatically on the philosophy of the institution than on current practices of other institutions of higher learning. Yet an enlightened institution is bound to be interested in comparative practices as one of many sources of data. The job of the School of Education seems to be to agree on those goals of education which are consonant with its basic philosophy, pay heed to what psychology and related disciplines have to offer, observe what practices are being successfully followed elsewhere, and then proceed with caution in the area of program revision.

BUREAU OF EDNL. & PSYL. RESEARCH : : LIBRARY
DKB
21 JAN 1966